

METHODS AND COMPOSITIONS FOR SCREENING FOR  
MODULATORS OF IgE SYNTHESIS, SECRETION AND  
SWITCH REARRANGEMENT

FIELD OF THE INVENTION

The invention relates to methods and compositions useful in screening for modulators of IgE synthesis, secretion and switch rearrangement.

BACKGROUND OF THE INVENTION

Immunoglobulins must bind to a vast array of foreign molecules and thus exist in many forms. The sequence of the variable (V) region of immunoglobulin molecules varies tremendously, conferring virtually unlimited capacity to bind antigens. The constant (C) region comes in five different varieties:  $\alpha$ ,  $\delta$ ,  $\epsilon$ ,  $\gamma$  and  $\mu$ , providing five different isotypes: IgA, IgD, IgE, IgG and IgM, each of which performs a different set of functions. B cells initially produce only IgM and IgD, and must be activated or induced to produce the other isoforms, such as IgE.

The course of IgE production starts with the activation of B cells. Upon activation with an antigen, B cells follow one of two differentiation pathways: they may differentiate directly into plasma cells, which are basically antibody-secreting factories, or they may give rise to germinal centers, specialized structures within lymphoid organs. In the latter, successive

rounds of mutation of the V region genes is followed by expression of the gene products on the cell surface, with selection of the cells on the basis of the affinity of the mutated immunoglobulins against the antigen.

5 In both pathways of antigen-induced B cell differentiation, isotype switching occurs in which the C region of the immunoglobulin heavy chain changes from the joint expression of IgM and IgD on naive B cells to expression of one of the downstream isotypes such as IgE. This switching involves the replacement of upstream C regions with a downstream C region that has biologically distinct effector functions without changing the structure of the variable portion and, hence, its specificity. For IgE switching, a deletional rearrangement of the Ig heavy chain gene locus occurs, a rearrangement that joins the switch region of the  $\mu$  gene, S $\mu$ , with the corresponding region of the  $\epsilon$  gene, S $\epsilon$ . This switching is minimally induced by IL-4 or IL-13, which initiates transcription through the S $\epsilon$  region, resulting in the synthesis of germ-line (or "sterile")  $\epsilon$  transcripts; that is, transcripts of the unrearranged C $\epsilon$  heavy genes. This IL-4 induced transcription is inhibited by IFN- $\gamma$ , IFN- $\alpha$ , and TGF- $\beta$ . A second signal, normally delivered by T cells, is required for actual switch recombination leading to IgE production. The T cell signal may be replaced by monoclonal antibodies to CD40, Epstein-Barr viral infection, or hydrocortisone.

20 Recently, the mechanism of class switch recombination has been explained by an accessibility model, wherein the specificity of the switch gene rearrangement is determined by the modulation of switch region accessibility; that is, the opening up of the chromatin in certain areas, allowing the required protein/enzyme complexes access to the genes.

25 IgE antibodies are crucial immune mediators of allergic reactions, and have been shown to be responsible for the induction and maintenance of allergic symptoms. For example, the introduction of anti-IgE antibodies has been shown to interfere with IgE function, thus working to alleviate allergic symptoms. See Jardieu, Current Op. Immunol. 7:779-782 (1995), Shields et al., Int. Arch. Allergy. Immunol. 107:308-312 (1995).

Accordingly, it is an object of the invention to provide compositions and methods useful in screening for modulators of IgE production, in particular for modulators of switch rearrangement.

5 SUMMARY OF THE INVENTION

10 In accordance with the objects outlined above, the present invention provides methods of screening for bioactive agents capable of inhibiting an IL-4 inducible  $\epsilon$  promoter. The method comprises combining a candidate bioactive agent and a cell comprising a fusion nucleic acid. The fusion nucleic acid comprises an IL-4 inducible  $\epsilon$  promoter, and a reporter gene. The promoter is then induced with IL-4 (or IL-13), and the presence or absence of the reporter protein is detected. Generally, the absence of the reporter protein indicates that the agent inhibits the IL-4 inducible  $\epsilon$  promoter. The fusion nucleic acid may comprise an exogenous IL-4 inducible  $\epsilon$  promoter, or an endogenous IL-4 inducible  $\epsilon$  promoter. Preferred embodiments utilize the use of retroviral vectors to introduce the candidate bioactive agents.

15 In an additional aspect, the present invention provides cell lines for screening. Either CA-46 and MC-116 cell lines are included, and further comprise fusion nucleic acids comprising an IL-4 inducible  $\epsilon$  promoter, and a reporter gene.

20 In a further aspect, the present invention provides methods of screening for bioactive agents capable of modulating IgE production. The method comprises combining a candidate bioactive agent and a cell capable of expressing IgE and determining the amount of IgE produced in the cell. Generally, a change in the amount of IgE as compared to the amount produced in the absence of the candidate agent indicates that the agent modulates IgE production. The cell can further comprise a IgE fusion protein comprises the  $\epsilon$  heavy chain, and a fluorescent protein.

In an additional aspect, the invention provides methods of screening for bioactive agents capable of inhibiting a promoter of interest. The method comprises combining a candidate bioactive agent and a cell comprising a fusion nucleic acid. The fusion nucleic acid comprises a promoter of interest and a reporter gene comprising a death gene that is activated by the introduction of a ligand. The promoter is optionally induced, and the ligand is introduced to the cell. The presence of the cell is then detected, wherein the presence of the cell indicates that the agent inhibits the promoter.

In a further aspect, the invention provides compositions comprising a test vector and a reporter vector. The test vector comprises a first selection gene, and a fusion gene comprising a first sequence encoding a transcriptional activation domain, and a second sequence encoding a test protein. The reporter vector comprises a first detectable gene, and all or part of the switch  $\epsilon$  sequence, which upon binding of the transcriptional activation domain due to a protein-nucleic acid interaction between the test protein and the switch  $\epsilon$  sequence, will activate transcription of the first detectable gene. Methods utilizing these compositions are also provided; the methods comprise providing a host cell comprising the composition, and subjecting the host cell to conditions under which the fusion gene is expressed to produce a fusion protein. A protein-nucleic acid interaction between the fusion protein and the switch  $\epsilon$  sequence is then detected.

#### BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A and 1B depict the germline  $\epsilon$  locus and sequence. Fig. 1A depicts the sequence of the human IL-4 inducible  $\epsilon$  promoter. Fig. 1B depicts the organization of the germline  $\epsilon$  locus.

Figures 2A and 2B depict the regions (2A) and sequences (2B and 2C) of the switch  $\epsilon$  (S $\epsilon$ ) region that are used in methods of screening for proteins that interact with the S $\epsilon$  region, as described below.

Figure 3 shows a schematic of the yeast one-hybrid system used to identify proteins that bind to the SE region.

Figure 4 depicts the IL-4 induction of germline  $\epsilon$  mRNA in three IgM<sup>+</sup> B cell lines, CA-46, MC-116 and DND39. The cells were incubated for 48 hours in 300 U/ml of hIL-4. RT-PCR was performed using primers specific for the germline  $\epsilon$  exon and the 5'-end of the  $\epsilon$  CH1 exon (predicted size is ~200 bp).

Figures 5A, 5B, 5C and 5D depict two general approaches to generate germline  $\epsilon$  promoter knock-in reporter cell lines. Figure 5A shows the organization of this region in vivo. Figures 5B and 5C depict two possible knock in constructs. The IL-4 inducible IgM<sup>+</sup> B cell lines are transfected with one or both of these constructs. Under the influence of IL-4, GFP and/or BFP positive clones are isolated by FACS. Homologous recombination can be confirmed by PCR and/or Southern blot hybridization. Figure 5D depicts an alternate construct. In this embodiment, the IL-4 inducible IgM<sup>+</sup> B cell lines are transfected with the 5D construct and selected with G418. Survivors are sorted for the lack of the 3' BFP expression (deleted during homologous recombination). RT-PCR is performed to confirm homologous recombination. Those clones are transfected with cre to remove the neomycin resistance gene.

Figure 6 depicts a preferred vector for introducing a peptide library into cell lines containing knock-in reporter genes under the control of the IL-4 inducible  $\epsilon$  promoter. CRU5 is a modified LTR; Naviaux, et al., "The pCL Vector System: Rapid Production of Helper-Free, High-Titer, Recombinant Retroviruses," *Journal of Virology*, 70(8):5701-5705 (1996); LTR = long terminal repeat;  $\Psi$ <sup>+</sup> = packaging signal; localization signal = nuclear, cell membrane, etc.; MCS = multiple cloning site; IRES = internal ribosome entry site; 2a = self-cleaving peptide. All the components are cassetted for flexibility.

Figure 7 depicts a general schematic of the generation of the primary peptide libraries in retroviruses.

Figures 8A and 8B depict constructs useful in generating  $\epsilon$  heavy chain knock-in cell lines. Figure 8A depicts the wild-type organization. Figure 8B depicts a representative construct to produce a GFP knock-in. S = secretory exon; GFP = green fluorescent protein; BFP = blue fluorescent protein; Neo<sup>r</sup> = neomycin resistance gene; VDJ = V region exon; CH1, 2, 3, 4 = constant region domain exons; M1, M2 = membrane exons; HSV-TK = Herpes Simplex Virus - thymidine kinase.

Figures 9A and 9B depict constructs useful in the invention. Figure 9A shows a reporter construct useful to create an IL-4 inducible  $\epsilon$  promoter reporter cell line. CRU5 = hCMV promoter plus R and U5 regions of LTR; BGH poly A = bovine growth hormone polyadenylation signal; SIN = self-inactivating LTR. Figure 9B shows a library construct.

Figures 10A and 10B depict a schematic of the screen for candidate agents of the germline  $\epsilon$  promoter. Figure 10A: the experimental schematic. Figure 10B depicts the survival construct useful in the screen. Position 1 can be a number of different genes, including a FAS chimeric receptor outlined herein (including extracellular mouse Fas receptor or mouse CD8 receptor coupled with the human transmembrane and cytoplasmic Fas receptor), HSV-TK, p450 2B1 and p21 peptide.

Figures 11A, 11B and 11C depict preferred vectors and their sequences.

Figures 12A, 12B and 12C depict a construct useful in the present invention, comprising the a Fas survival construct (i.e. the use of a death gene). The sequence is of the inducible  $\epsilon$  promoter-chimeric Fas-IRES-hygromycin-bovine growth hormone poly A tail that is put into the C12s vector backwards to that no leaky transcription happens through the cmv promoter.

Figures 13A, 13B and 13C depict a construct useful in the present invention, comprising the a Fas survival construct (i.e. the use of a death gene). The sequence is of the inducible  $\epsilon$  promoter-chimeric Fas (either CD8 or mLyt2)-IRES-hygromycin-bovine growth hormone

poly A tail that is put into the C12s vector backwards to that no leaky transcription happens through the cmv promoter.

## DETAILED DESCRIPTION OF THE INVENTION

5

The present invention provides compositions and methods useful in screening for modulators, particularly inhibitors, of the production of IgE antibodies. In particular, assay methodologies are provided that are amenable to high-throughput screening strategies, such that large numbers of potential drugs may be screened rapidly and efficiently. Generally, traditional treatments for IgE suppression are based on regulation of the system after IgE has been made, for example using anti-IgE antibodies or anti-histamines, to modulate the IgE-mediated response resulting in mast cell degranulation. In some cases, drugs are known that generally downregulate IgE production or that inhibit switching but not induction of germline transcripts (see for example Loh et al., J. Allerg. Clin. Immunol. 97(5):1141 (1996)).

10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95  
100  
105  
110  
115  
120  
125  
130  
135  
140  
145  
150  
155  
160  
165  
170  
175  
180  
185  
190  
195  
200  
205  
210  
215  
220  
225  
230  
235  
240  
245  
250  
255  
260  
265  
270  
275  
280  
285  
290  
295  
300  
305  
310  
315  
320  
325  
330  
335  
340  
345  
350  
355  
360  
365  
370  
375  
380  
385  
390  
395  
400  
405  
410  
415  
420  
425  
430  
435  
440  
445  
450  
455  
460  
465  
470  
475  
480  
485  
490  
495  
500  
505  
510  
515  
520  
525  
530  
535  
540  
545  
550  
555  
560  
565  
570  
575  
580  
585  
590  
595  
600  
605  
610  
615  
620  
625  
630  
635  
640  
645  
650  
655  
660  
665  
670  
675  
680  
685  
690  
695  
700  
705  
710  
715  
720  
725  
730  
735  
740  
745  
750  
755  
760  
765  
770  
775  
780  
785  
790  
795  
800  
805  
810  
815  
820  
825  
830  
835  
840  
845  
850  
855  
860  
865  
870  
875  
880  
885  
890  
895  
900  
905  
910  
915  
920  
925  
930  
935  
940  
945  
950  
955  
960  
965  
970  
975  
980  
985  
990  
995

In contrast, the present invention provides several related techniques that may be used to screen for upstream modulators of IgE production, to prevent the production of IgE and thus reduce or eliminate the allergic response. For example, an early step in the Ig switch is the production of sterile  $\epsilon$  transcripts in response to IL-4. It is also appreciated that blockage of the production of membrane bound IgE may induce programmed cell death (PCD). By interfering at this step, highly efficient, rapid and prolonged inhibition of the allergic response may occur. In addition, these techniques allow individual cell assessment and thus are useful for high-throughput screening strategies, for example those that utilize fluorescence activated cell sorting (FACS) techniques, and thus allow screening of large numbers of compounds for their effects on IgE production.

In a preferred embodiment, the invention relates to methods that rely on reporter genes fused to IgE promoters, such as the IL-4 inducible  $\epsilon$  promoter that starts a cascade that ultimately results in IgE production. Using novel reporter constructs, screening for modulators of this

promoter system may be done. Thus the invention provides a number of different constructs that allow for screening for antagonists and agonists of these promoters.

In a preferred embodiment, the invention provides methods of screening for bioactive agents capable of modulating, particularly inhibiting, an IL-4 inducible promoter. By "an IL-4 inducible promoter" herein is meant a nucleic acid promoter that is induced by IL-4, putatively by binding an unknown IL-4 induced DNA binding protein that results in induction of the promoter; that is, the introduction of IL-4 causes the pronounced activation of a particular DNA binding protein that then binds to the IL-4 inducible promoter segment and induces transcription. The sequence of the human IL-4 inducible promoter is shown in Figure 1, and as will be appreciated by those in the art, derivatives or mutant promoters are included within this definition. Particularly included within the definition of an IL-4 inducible promoter are fragments or deletions of the sequence shown in Figure 1. As is known in the art, the IL-4 inducible promoter is also inducible by IL-13. By "modulating an IL-4 inducible promoter" herein is meant either an increase or a decrease (inhibition) of promoter activity, for example as measured by the presence or quantification of transcripts or of translation products. By "inhibiting an IL-4 inducible promoter" herein is meant a decrease in promoter activity, with changes of at least about 50% being preferred, and at least about 90% being particularly preferred.

The methods comprise combining a candidate bioactive agent and a cell or a population of cells comprising a fusion nucleic acid. The cell or cells comprise a fusion nucleic acid. In a preferred embodiment, the fusion nucleic acid comprises an IL-4 inducible promoter and at least a first reporter gene. The IL-4 inducible promoter is as described herein, for example SEQ ID NO:1, or derivatives thereof, and may be either an endogenous or exogenous IL-4 inducible promoter, as is more fully described below.

By "reporter gene" or "selection gene" herein is meant a gene that by its presence in a cell (i.e. upon expression) can allow the cell to be distinguished from a cell that does not contain the reporter gene. Reporter genes can be classified into several different types, including



detection genes, survival genes, death genes and cell cycle genes. It may be the nucleic acid or the protein expression product that causes the effect. As is more fully outlined below, additional components, such as substrates, ligands, etc., may be additionally added to allow selection or sorting on the basis of the reporter gene.

5

In a preferred embodiment, the reporter gene encodes a protein that can be used as a direct label, i.e. a detection gene, for sorting the cells, i.e. for cell enrichment by FACS. In this embodiment, the protein product of the reporter gene itself can serve to distinguish cells that are expressing the reporter gene. In this embodiment, suitable reporter genes include those encoding green fluorescent protein (GFP; Chalfie, et al., "Green Fluorescent Protein as a Marker for Gene Expression," *Science* 263(5148):802-805 (Feb 11, 1994); and EGFP; Clontech - Genbank Accession Number U55762 ), blue fluorescent protein (BFP; 1. Quantum Biotechnologies, Inc. 1801 de Maisonneuve Blvd. West, 8th Floor, Montreal (Quebec) Canada H3H 1J9; 2. Stauber, R. H. *Biotechniques* 24(3):462-471 (1998); 3. Heim, R. and Tsien, R. Y. *Curr. Biol.* 6:178-182 (1996)), enhanced yellow fluorescent protein (EYFP; 1. Clontech Laboratories, Inc., 1020 East Meadow Circle, Palo Alto, CA 94303), luciferase (Ichiki, et al.), and  $\beta$ -galactosidase (Nolan, et al., "Fluorescence-Activated Cell Analysis and Sorting of Viable Mammalian Cells Based on Beta-D-galactosidase Activity After Transduction of Escherichia Coli LacZ," *Proc Natl Acad Sci USA* 85(8):2603-2607 (Apr 1988)).

Alternatively, the reporter gene encodes a protein that will bind a label that can be used as the basis of the cell enrichment (sorting); i.e. the reporter gene serves as an indirect label or detection gene. In this embodiment, the reporter gene should encode a cell-surface protein.

For example, the reporter gene may be any cell-surface protein not normally expressed on the surface of the cell, such that secondary binding agents could serve to distinguish cells that contain the reporter gene from those that do not. Alternatively, albeit non-preferably, reporters comprising normally expressed cell-surface proteins could be used, and differences between cells containing the reporter construct and those without could be determined. Thus, secondary binding agents bind to the reporter protein. These secondary binding agents are

preferably labelled, for example with fluors, and can be antibodies, haptens, etc. For example, fluorescently labeled antibodies to the reporter gene can be used as the label. Similarly, membrane-tethered streptavidin could serve as a reporter gene, and fluorescently-labeled biotin could be used as the label, i.e. the secondary binding agent. Alternatively, the secondary binding agents need not be labeled as long as the secondary binding agent can be used to distinguish the cells containing the construct; for example, the secondary binding agents may be used in a column, and the cells passed through, such that the expression of the reporter gene results in the cell being bound to the column, and a lack of the reporter gene (i.e. inhibition), results in the cells not being retained on the column. Other suitable reporter proteins/secondary labels include, but are not limited to, antigens and antibodies, enzymes and substrates (or inhibitors), etc.

In a preferred embodiment, the reporter gene is a survival gene that serves to provide a nucleic acid (or encode a protein) without which the cell cannot survive, such as drug resistant genes. In this embodiment, the assays may rely on clonal or pooled populations of cells, since if inhibitors of the promoter are found, the cells will die, necessitating a clonal population in order to determine the candidate agent.

In a preferred embodiment, the reporter gene is a cell cycle gene, that is, a gene that causes alterations in the cell cycle. For example, p21 protein its ligand (a collection of three proteins; see Harper, et al., "The p21 Cdk-Interacting Protein Cip1 Is a Potent Inhibitor of G1 Cyclin-Dependent Kinases," *Cell* 75:805-816 (November 19, 1993)), which does not cause death, but causes cell-cycle arrest, such that cells containing inhibited IL-4 inducible promoters grow out much more quickly, allowing detection on this basis. As will be appreciated by those in the art, it is also possible to configure the system such that the cells containing the inhibited IL-4 inducible promoters do not grow out, and thus can be selected on this basis as well.

In a preferred embodiment, the reporter gene is a death gene that provides a nucleic acid that encodes a protein that causes the cells to die. Death genes fall into two basic categories:

death genes that encode death proteins that require a death ligand to kill the cells, and death genes that encode death proteins that kill cells as a result of high expression within the cell, and do not require the addition of any death ligand. It is preferable that cell death requires a two-step process: the expression of the death gene and induction of the death phenotype with a signal or ligand, such that the cells may be grown up expressing the death gene, and then induced to die. A number of death genes/ligand pairs are known, including, but not limited to, the Fas receptor and Fas ligand (Bodmer, et al., "Characterization of Fas," *J Biol Chem* 272(30):18827-18833 (Jul 25, 1997); muFAS, Gonzalez-Cuadrado, et al., "Agonistic anti-Fas Antibodies Induce Glomerular Cell Apoptosis in Mice In Vivo," *Kidney Int* 51(6):1739-1746 (Jun 1997); Muruva, et al., *Hum Gene Ther*, 8(8):955 (May 1997)), (or anti-Fas receptor antibodies); p450 and cyclophosphamide (Chen, et al., "Potentiation of Cytochrome P450/Cyclophosphamide-Based Cancer Gene Therapy By Coexpression of the P450 Reductase Gene," *Cancer Res* 57(21):4830-4837 (Nov 1 1997)); thymidine kinase and gangcylovir (Stone, R., "Molecular 'Surgery' For Brain Tumors," 256(5063):1513 (June 12, 1992)), tumor necrosis factor (TNF) receptor and TNF. Alternatively, the death gene need not require a ligand, and death results from high expression of the gene; for example, the overexpression of a number of programmed cell death (PCD) proteins are known to cause cell death, including, but not limited to, caspases, bax, TRADD, FADD, SCK, MEK, etc.

As will be appreciated by those in the art, the use of the death genes in the manner described herein, particularly in two-step applications, allows general and high-throughput screening for inhibitors of other promoters, in addition to the IL-4 inducible  $\epsilon$  promoters described herein. Thus, the present invention provides fusion nucleic acids comprising a promoter of interest operably linked to a death gene for use in screening methods. The promoter of interest can be either a constitutive promoter or an inducible promoter, such as the IL-4 inducible  $\epsilon$  promoter. As will be appreciated by those in the art, any number of possible promoters could be used. Suitable promoters of interest include, but are not limited to, inducible promoters such as IL-4  $\epsilon$  promoter, promoters that are induced by cytokines or growth factors such as the interferon responsive factors 1 to 4, NFkB (Fiering, et al., "Single Cell Assay of a Transcription Factor Reveals a Threshold in Transcription Activated By Signals Emanating

From the T-Cell Antigen Receptor,” *Genes Dev* 4(10):1823-1834 (Oct 1990)), etc. When inducible promoters are used in this embodiment, suitable cell types are those that can be induced by the appropriate inducer, as will be appreciated by those in the art.

5 Preferred embodiments fall into one of three configurations. In a preferred embodiment, the promoter of interest is a constitutive promoter, and it is hooked to a death gene that requires the presence of a ligand, such as Fas or TNF. Thus, the cells can be grown up and the presence of the death gene verified due to the constitutive promoter. This is generally done by hooking the death gene up to a detection gene such as GFP or BFP, etc., using either an IRES or a protease cleavage site as is outlined below; thus, the presence of the detection gene means the death gene is also present. Verification of the presence of the death gene is preferred to keep the levels of false positives low; that is, cells that survive the screen should be due to the presence of an inhibitor of the promoter rather than a lack of the death gene.

10 Once the cells have been enriched for those containing the death gene, the candidate agents can be added (and their presence verified as well), followed by induction in the presence of IL-4, and finally by addition of the death ligand. Thus, the cell population is enriched for those cells that have an agent that inhibits the promoter and thus does not produce the death protein, i.e. those that survive.

15 Alternatively, a preferred embodiment utilizes fusion nucleic acids comprising promoters of interest that are inducible (such as the IL-4  $\epsilon$  promoter), and hooked to a death gene that requires a death ligand. The presence of the death gene is verified by inducing the promoter, causing the death gene (and preferably a detection gene) to be made. The candidate agents and death ligands are then introduced in the presence of their appropriate inducer, and the population is enriched for those cells that survive, i.e. contain an agent that inhibits the promoter and thus does not produce the death protein.

20 When death genes that require ligands are used, i.e. for “two step” processes, preferred embodiments utilize chimeric death genes, i.e. chimeric death receptor genes. These chimeric

death receptors comprise the extracellular domain of a ligand-activated multimerizing receptor and the endogeneous cytosolic domain of a death receptor gene, such as Fas or TNF. This is done to avoid endogeneous activation of the death gene. The mechanism of Fas-induced cell death involves the introduction of the Fas ligand, which can bind two monomeric Fas receptors, causing the multimerization of the receptor, which activates the receptor and leads to secondary signalling resulting in caspase activation and PCD. However, as will be appreciated by those in the art, it is possible to substitute the extracellular portion of the death receptor with the extracellular portion of another ligand-activated multimerizing receptor, such that a completely different signal activates the cell to die. There are a number of known ligand-activated dimerizing receptors, including, but not limited to, the CD8 receptor, erythropoietin receptor, thrombopoietin receptor, growth hormone receptor, Fas receptor, platelet derived growth hormone receptor, epidermal growth factor receptor, leptin receptor, and a variety of interleukin receptors (including, but not limited to, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-11, IL-12, IL-13, IL-15, and IL-17; although the use of the IL-4 and IL-13 receptors are not preferred, since these can be used to induce the promoter and thus does not provide a "two step" death process), low-density lipoprotein receptor, prolactin receptor, and transferrin receptor.

In a preferred embodiment, chimeric Fas receptor genes are made. The exact combination will depend on the cell type used and the receptors normally produced by these cells. For example, when using human cells or cell lines, a non-human extracellular domain and a human cytosolic domain are preferred, to prevent endogeneous induction of the death gene. For example, a preferred embodiment utilizes human cells, a murine extracellular Fas receptor domain and a human cytosolic domain, such that the endogeneous human Fas ligand will not activate the murine domain. Alternatively, human extracellular domains may be used when the cells used do not endogeneously produce the ligand; for example, the human EPO extracellular domain may be used when the cells do not endogeneously produce EPO. (Kawaguchi, et al., *Cancer Lett.*, 116(1):53 (1997); Takebayashi, et al., *Cancer Res.*, 56(18):4164 (1996); Rudert, et al., *Biochem Biophys Res Commun.*, 204(3):1102 (1994); Rudert, et al., *DNA Cell Biol.*, 16(2):197 (1997); Takahasi, et al., *J Biol Chem.*

271(29):17555 (1996); Adam, et al., *J Biol Chem.*, 268(26):19882 (1993); Mares, et al.,  
*Growth Factors*, 6(2):93 (1992); Seedorf, et al., *J Biol Chem.*, 266(19):12424 (1991);  
Heidaran, et al., *J Biol Chem.*, 265(31):18741 (1990); Okuda, et al., *J Clin Invest.*  
100(7):1708 (1997); Allgood, et al., *Curr Opin Biotechnol.*, 8(4):474 (1997); Anders, et al., *J*  
5 *Biol Chem.*, 271(36):21758 (1996); Krishnan, et al., *Oncogene*, 13(1):125 (1996); Declercq,  
et al., *Cytokine*, 7(7):701 (1995); Bazzoni, et al., *Proc Natl Acad Sci U S .*, 92(12):5380  
(1995); Ohashi, et al., *Proc Natl Acad Sci U S A* , 91(1):158 (1994); Desai, et al., *Cell*,  
73(3):541 (1993); and Amara, et al., *Proc Natl Acad Sci U S A*, 94(20):10618 (1997)).

10 In addition to the extracellular domain and the cytosolic domain, these receptors have a  
transmembrane domain. As will be appreciated by those in the art, for chimeric death  
receptor genes, the transmembrane domain from any of the receptors can be used, although in  
general, it is preferred to use the transmembrane domain associated with the chosen cytosolic  
domain, to preserve the interaction of the transmembrane domain with other endogenous  
15 signalling proteins.

Thus, preferred embodiments provide fusion nucleic acids that utilize the IL-4 inducible  $\epsilon$   
promoter linked to a death gene, particularly a chimeric death receptor gene, that requires a  
death ligand for cell killing.

20 Alternatively, inducible promoters can be linked to "one step" death genes, i.e. death genes  
that upon a certain threshold expression, will kill a cell without requiring a ligand or  
secondary signal. In this embodiment, the inducible promoter is preferably "leaky", such that  
some small amount of death gene and a required secondary reporter gene such as a survival  
25 gene or a detection gene can be expressed. The cells that contain the death gene can then be  
selected on this basis, to avoid false positives. Once the presence of the construct is verified,  
candidate agents are added (and their presence preferably verified, using a detection or  
selection gene as well), and the promoter is induced. The population is then enriched for  
those cells that contain agents that inhibit the promoter, i.e. that will survive.

In a preferred embodiment, additional reporter genes are used, particularly when inducible death genes are used. In a preferred embodiment, the additional reporter gene is a selection gene. The cells containing the death gene and the drug selectable gene are grown; if the appropriate drug is added to the culture, only those cells containing the resistance gene (and hence the death gene) survive. This ensures that the cells are expressing the death gene to decrease "false positives", i.e. cells that do not die because they do not contain the death gene.

In an additional preferred embodiment, the additional reporter gene is a labeling gene such as GFP. The use of a detection gene allows cells to be sorted to give a population enriched for those containing the construct. As outlined above, a preferred embodiment uses "leaky" inducible promoters; that is, the cells are selected such that the IL-4 inducible promoter, even in the absence of IL-4 or IL-13, produces some GFP and death gene (for example, the Fas receptor constructs). In this embodiment, suitably "leaky" promoters are chosen such that some GFP is expressed (preferably enough to select the cells expressing the construct from those that are not), but not enough death gene is produced to cause death. While preferred embodiments utilize death genes requiring the addition of a death ligand, it is well known that high levels of some death genes, even in the absence of death ligand, can cause death. Thus, for example, high levels of Fas receptor expression can cause multimerization, and thus activation, even in the absence of the Fas ligand.

In a preferred embodiment, when two reporter genes are used, they are fused together in such a way as to only require a single promoter, and thus some way of functionally separating the two genes is preferred. This can be done on the RNA level or the protein level. Preferred embodiments utilize either IRES sites (which allows the translation of two different genes on a single transcript (Kim, et al., "Construction of a Bifunctional mRNA in the Mouse By Using the Internal Ribosomal Entry Site of the Encephalomyocarditis Virus," *Molecular and Cellular Biology* 12(8):3636-3643 (Aug 1992) and McBratney, et al., "The Sequence Context of the Initiation Codon in the Encephalomyocarditis Virus Leader Modulates Efficiency of Internal Translation Initiation," *Current Opinion in Cell Biology* 5:961-965 (1993)), or a protease cleavage site (which cleaves a protein translation product into two proteins).

Preferred protease cleavage sites include, but are not limited to, the 2a site (Ryan et al., J. Gen. Virol. 72:2727 (1991); Ryan et al., EMBO J. 13:928 (1994); Donnelly et al., J. Gen. Virol. 78:13 (1997); Hellen et al., *Biochem*, 28(26):9881 (1989); and Mattion et al., J. Virol. 70:8124 (1996), all of which are expressly incorporated by reference), prosequences of retroviral proteases including human immunodeficiency virus protease and sequences recognized and cleaved by trypsin (EP 578472, Takasuga et al., J. Biochem. 112(5):652 (1992)) factor X<sub>a</sub> (Gardella et al., J. Biol. Chem. 265(26):15854 (1990), WO 9006370), collagenase (J03280893, Tajima et al., J. Ferment. Bioeng. 72(5):362 (1991), WO 9006370), clostripain (EP 578472), subtilisin (including mutant H64A subtilisin, Forsberg et al., J. Protein Chem. 10(5):517 (1991), chymosin, yeast KEX2 protease (Bourbonnais et al., J. Bio. Chem. 263(30):15342 (1988), thrombin (Forsberg et al., supra; Abath et al., BioTechniques 10(2):178 (1991)), *Staphylococcus aureus* V8 protease or similar endoproteinase-Glu-C to cleave after Glu residues (EP 578472, Ishizaki et al., Appl. Microbiol. Biotechnol. 36(4):483 (1992)), cleavage by NIa proteainase of tobacco etch virus (Parks et al., Anal. Biochem. 216(2):413 (1994)), endoproteinase-Lys-C (U.S. Patent No. 4,414,332) and endoproteinase-Asp-N, *Neisseria* type 2 IgA protease (Pohlner et al., Bio/Technology 10(7):799-804 (1992)), soluble yeast endoproteinase yscF (EP 467839), chymotrypsin (Altman et al., Protein Eng. 4(5):593 (1991)), enteropeptidase (WO 9006370), lysostaphin, a polyglycine specific endoproteinase (EP 316748), and the like. See e.g. Marston, F.A.O. (1986) Biol. Chem. J. **240**, 1-12.

In addition to the promoter of interest, such as an IL-4 inducible  $\epsilon$  promoter and reporter gene, the fusion nucleic acids may comprise additional components, including, but not limited to, other reporter genes, protein cleavage sites, internal ribosome entry (IRES) sites, AP-1 sites, and other components as will be appreciated by those in the art.

In a preferred embodiment, foreign constructs comprising the IL-4 inducible  $\epsilon$  promoter and the reporter gene are made. By "foreign" herein is meant that the fusion nucleic acids originates outside of the cells. That is, a recombinant nucleic acid is made that contains an exogenous IL-4 inducible  $\epsilon$  promoter and a reporter gene. Thus, in some circumstances, the



cells will contain both exogenous and endogenous IL-4 inducible  $\epsilon$  promoters. By “recombinant nucleic acid” herein is meant nucleic acid, originally formed *in vitro*, in general, by the manipulation of nucleic acid by endonucleases, in a form not normally found in nature. Thus an isolated nucleic acid, in a linear form, a nucleic acid containing components not normally joined, such as an IL-4 inducible promoter and a reporter gene, or an expression vector formed *in vitro* by ligating DNA molecules that are not normally joined, are all considered recombinant for the purposes of this invention. It is understood that once a recombinant nucleic acid is made and reintroduced into a host cell or organism, it will replicate non-recombinantly, i.e. using the in vivo cellular machinery of the host cell rather than *in vitro* manipulations; however, such nucleic acids, once produced recombinantly, although subsequently replicated non-recombinantly, are still considered recombinant for the purposes of the invention. In this embodiment, any cells that express an IL-4 receptor that transduces the IL-4 signal to the nucleus and alters transcription can be used. Suitable cells include, but are not limited to, human cells and cell lines that show IL-4/13 inducible production of germline  $\epsilon$  transcripts, including, but not limited to, DND39 (see Watanabe, supra), MC-116, (Kumar, et al., “Human BCGF-12kD Functions as an Autocrine Growth Factor in Transformed B Cells,” *Eur Cytokine Netw* 1(2):109 (1990)), CA-46 (Wang, et al., “UCN-01: A Potent Abrogator of G2 Checkpoint Function in Cancer Cells with Dirupted p53,” *J Natl Cancer Inst* 88:956 (1996)).

This recombinant nucleic acid may introduced to a cell in a variety of ways, as will be appreciated by those in the art, including, but not limited to,  $\text{CaPO}_4$  precipitation, liposome fusion, lipofectin®, electroporation, viral infection, etc. The constructs may preferably stably integrate into the genome of the host cell (for example, with retroviral introduction, outlined below), or may exist either transiently or stably in the cytoplasm (i.e. through the use of traditional plasmids, utilizing standard regulatory sequences, selection markers, etc.).

In a preferred embodiment, the exogeneous constructs, which may be in the form of an expression vector, are added as retroviral constructs, using techniques generally described in

PCT US97/01019 and PCT US97/01048, both of which are expressly incorporated by reference, and the examples.

In a preferred embodiment, the fusion construct comprises an endogenous IL-4 inducible  $\epsilon$  promoter and an exogenous reporter gene; "endogenous" in this context means originating within the cell. That is, gene "knock-in" constructions are made, whereby an exogenous reporter gene as outlined herein is added, via homologous recombination, to the genome, such that the reporter gene is under the control of the endogenous IL-4 inducible  $\epsilon$  promoter.

This may be desirable to allow for the exploration and modulation of the full range of endogenous regulation, i.e. regulatory elements (particularly those flanking the promoter) other than just the IL-4 inducible  $\epsilon$  promoter fragment. Exemplary constructs are shown in Figures 5B and 5C, with GFP and BFP, although other reporter genes outlined herein may be used.

Homologous recombination may proceed in several ways. In one embodiment, traditional homologous recombination is done, with molecular biological techniques such as PCR being done to find the correct insertions. For example, gene "knock-ins" may be done as is known in the art, for example see Westphal et al., Current Biology 7:R530-R533 (1997), and references cited therein, all of which are expressly incorporated by reference. The use of recA mediated systems may also be done, see PCT US93/03868, hereby expressly incorporated by reference.

Alternatively, and preferably, the selection of the "knock ins" are done by FACS on the basis of the incorporation of a reporter gene. Thus, in a preferred embodiment, a first homologous recombination event is done to put a first reporter gene, such as GFP, into at least one allele of the cell genome. Preferably, this is a cell type that exhibits IL-4 inducible production of at least germline  $\epsilon$  transcripts, so that the cells may be tested by IL-4 production for reporter gene expression. Suitable cells include, but are not limited to, human cells and cell lines that show IL-4/13 inducible production of germline  $\epsilon$  transcripts, including, but not limited to, DND39 (see Watanabe, supra), MC-116, (Kumar, et al., "Human BCGF-12kD Functions as

an autocrine Growth Factor in Transformed B Cells,” *Eur Cytokine Netw* 1(2):109 (1990)), CA-46 (Wang, et al., “UCN-01: A Potent Abrogator of G2 Checkpoint Function in Cancer Cells with Disrupted p53,” *J Natl Cancer Inst* 88:956 (1996)). As is noted herein, the ability of MC-116 and CA-46 cells to produce germline  $\epsilon$  transcripts upon IL-4/13 induction was not known prior to the present invention. Thus, preferred embodiments provide MC-116 and/or CA-46 cells comprising recombinant nucleic acid reporter constructs are outlined herein.

In a preferred embodiment, once a first endogenous promoter has been combined with an exogenous reporter construct, a second homologous recombination event may be done, preferably using a second reporter gene different from the first, such as BFP, to target the other allele of the cell genome, and tested as above.

Generally, IL-4 induction of the reporter genes will indicate the correct placement of the genes, which can be confirmed via sequencing such as PCR sequencing or Southern blot hybridization. In addition, preferred embodiments utilize prescreening steps to remove “leaky” cells, i.e. those showing constitutive expression of the reporter gene.

Thus, in a preferred embodiment, the invention provides cell lines that contain fusion nucleic acids comprising IL-4 inducible  $\epsilon$  promoter operably connected to at least one reporter gene. Once made, the cell lines comprising these reporter constructs are used to screen candidate bioactive agents for the ability to modulate the production of IgE, as is outlined below.

The term “candidate bioactive agent” or “exogenous compound” as used herein describes any molecule, *e.g.*, protein, oligopeptide, small organic molecule, polysaccharide, polynucleotide. Generally a plurality of assay mixtures are run in parallel with different agent concentrations to obtain a differential response to the various concentrations. Typically, one of these concentrations serves as a negative control, *i.e.*, at zero concentration or below the level of detection.

Candidate agents encompass numerous chemical classes, though typically they are organic molecules, preferably small organic compounds having a molecular weight of more than 100 and less than about 2,500 daltons. Candidate agents comprise functional groups necessary for structural interaction with proteins, particularly hydrogen bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups. The candidate agents often comprise cyclical carbon or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more of the above functional groups. Candidate agents are also found among biomolecules including peptides, saccharides, fatty acids, steroids, purines, pyrimidines, derivatives, structural analogs or combinations thereof. Particularly preferred are peptides.

Candidate agents are obtained from a wide variety of sources including libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available or readily produced. Additionally, natural or synthetically produced libraries and compounds are readily modified through conventional chemical, physical and biochemical means. Known pharmacological agents may be subjected to directed or random chemical modifications, such as acylation, alkylation, esterification, amidification to produce structural analogs.

In a preferred embodiment, the candidate bioactive agents are proteins. By "protein" herein is meant at least two covalently attached amino acids, which includes proteins, polypeptides, oligopeptides and peptides. The protein may be made up of naturally occurring amino acids and peptide bonds, or synthetic peptidomimetic structures. Thus "amino acid", or "peptide residue", as used herein means both naturally occurring and synthetic amino acids. For example, homo-phenylalanine, citrulline and noreleucine are considered amino acids for the purposes of the invention. "Amino acid" also includes imino acid residues such as proline and hydroxyproline. The side chains may be in either the (R) or the (S) configuration. In the preferred embodiment, the amino acids are in the (S) or L-configuration. If non-naturally

occurring side chains are used, non-amino acid substituents may be used, for example to prevent or retard *in vivo* degradations.

In a preferred embodiment, the candidate bioactive agents are naturally occurring proteins or fragments of naturally occurring proteins. Thus, for example, cellular extracts containing proteins, or random or directed digests of proteinaceous cellular extracts, may be used. In this way libraries of procaryotic and eucaryotic proteins may be made for screening in the systems described herein. Particularly preferred in this embodiment are libraries of bacterial, fungal, viral, and mammalian proteins, with the latter being preferred, and human proteins being especially preferred.

In a preferred embodiment, the candidate bioactive agents are peptides of from about 5 to about 30 amino acids, with from about 5 to about 20 amino acids being preferred, and from about 7 to about 15 being particularly preferred. The peptides may be digests of naturally occurring proteins as is outlined above, random peptides, or "biased" random peptides. By "randomized" or grammatical equivalents herein is meant that each nucleic acid and peptide consists of essentially random nucleotides and amino acids, respectively. Since generally these random peptides (or nucleic acids, discussed below) are chemically synthesized, they may incorporate any nucleotide or amino acid at any position. The synthetic process can be designed to generate randomized proteins or nucleic acids, to allow the formation of all or most of the possible combinations over the length of the sequence, thus forming a library of randomized candidate bioactive proteinaceous agents.

In one embodiment, the library is fully randomized, with no sequence preferences or constants at any position. In a preferred embodiment, the library is biased. That is, some positions within the sequence are either held constant, or are selected from a limited number of possibilities. For example, in a preferred embodiment, the nucleotides or amino acid residues are randomized within a defined class, for example, of hydrophobic amino acids, hydrophilic residues, sterically biased (either small or large) residues, towards the creation of

cysteines, for cross-linking, prolines for SH-3 domains, serines, threonines, tyrosines or histidines for phosphorylation sites, etc., or to purines, etc.

In a preferred embodiment, the candidate bioactive agents are nucleic acids. By "nucleic acid" or "oligonucleotide" or grammatical equivalents herein means at least two nucleotides covalently linked together. A nucleic acid of the present invention will generally contain phosphodiester bonds, although in some cases, as outlined below, nucleic acid analogs are included that may have alternate backbones, comprising, for example, phosphoramidate (Beaucage, *et al.*, Tetrahedron, **49**(10):1925 (1993) and references therein; Letsinger, *J. Org. Chem.*, **35**:3800 (1970); Sprinzl, *et al.*, Eur. J. Biochem., **81**:579 (1977); Letsinger, *et al.*, Nucl. Acids Res., **14**:3487 (1986); Sawai, *et al.*, Chem. Lett., 805 (1984), Letsinger, *et al.*, J. Am. Chem. Soc., **110**:4470 (1988); and Pauwels, *et al.*, Chemica Scripta, **26**:141 (1986)), phosphorothioate (Mag, *et al.*, Nucleic Acids Res., **19**:1437 (1991); and U.S. Patent No. 5,644,048), phosphorodithioate (Briu, *et al.*, J. Am. Chem. Soc., **111**:2321 (1989)), O-methylphosphoroamidite linkages (see Eckstein, Oligonucleotides and Analogues: A Practical Approach, Oxford University Press), and peptide nucleic acid backbones and linkages (see Egholm, J. Am. Chem. Soc., **114**:1895 (1992); Meier, *et al.*, Chem. Int. Ed. Engl., **31**:1008 (1992); Nielsen, Nature, **365**:566 (1993); Carlsson, *et al.*, Nature, **380**:207 (1996), all of which are incorporated by reference)). Other analog nucleic acids include those with positive backbones (Denpcy, *et al.*, Proc. Natl. Acad. Sci. USA, **92**:6097 (1995)); non-ionic backbones (U.S. Patent Nos. 5,386,023; 5,637,684; 5,602,240; 5,216,141; and 4,469,863; Kiedrowshi, *et al.*, Angew. Chem. Intl. Ed. English, **30**:423 (1991); Letsinger, *et al.*, J. Am. Chem. Soc., **110**:4470 (1988); Letsinger, *et al.*, Nucleoside & Nucleotide, **13**:1597 (1994); Chapters 2 and 3, ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research", Ed. Y.S. Sanghui and P. Dan Cook; Mesmaeker, *et al.*, Bioorganic & Medicinal Chem. Lett., **4**:395 (1994); Jeffs, *et al.*, J. Biomolecular NMR, **34**:17 (1994); Tetrahedron Lett., **37**:743 (1996)) and non-ribose backbones, including those described in U.S. Patent Nos. 5,235,033 and 5,034,506, and Chapters 6 and 7, ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research", Ed. Y.S. Sanghui and P. Dan Cook. Nucleic acids containing one or more carbocyclic sugars are also included within the

definition of nucleic acids (see Jenkins, *et al.*, Chem. Soc. Rev., (1995) pp. 169-176). Several nucleic acid analogs are described in Rawls, C & E News, June 2, 1997, page 35. All of these references are hereby expressly incorporated by reference. These modifications of the ribose-phosphate backbone may be done to facilitate the addition of additional moieties such as labels, or to increase the stability and half-life of such molecules in physiological environments. In addition, mixtures of naturally occurring nucleic acids and analogs can be made. Alternatively, mixtures of different nucleic acid analogs, and mixtures of naturally occurring nucleic acids and analogs may be made. The nucleic acids may be single stranded or double stranded, as specified, or contain portions of both double stranded or single stranded sequence. The nucleic acid may be DNA, both genomic and cDNA, RNA or a hybrid, where the nucleic acid contains any combination of deoxyribo- and ribo-nucleotides, and any combination of bases, including uracil, adenine, thymine, cytosine, guanine, inosine, xanthine hypoxanthine, isocytosine, isoguanine, etc.

As described above generally for proteins, nucleic acid candidate bioactive agents may be naturally occurring nucleic acids, random nucleic acids, or "biased" random nucleic acids. For example, digests of procaryotic or eucaryotic genomes may be used as is outlined above for proteins.

In a preferred embodiment, the candidate bioactive agents are organic chemical moieties, a wide variety of which are available in the literature.

In a preferred embodiment, a library of different candidate bioactive agents are used. Preferably, the library should provide a sufficiently structurally diverse population of randomized agents to effect a probabilistically sufficient range of diversity to allow binding to a particular target. Accordingly, an interaction library should be large enough so that at least one of its members will have a structure that gives it affinity for the target. Although it is difficult to gauge the required absolute size of an interaction library, nature provides a hint with the immune response: a diversity of  $10^7$ - $10^8$  different antibodies provides at least one combination with sufficient affinity to interact with most potential antigens faced by an

organism. Published *in vitro* selection techniques have also shown that a library size of  $10^7$  to  $10^8$  is sufficient to find structures with affinity for the target. A library of all combinations of a peptide 7 to 20 amino acids in length, such as generally proposed herein, has the potential to code for  $20^7$  ( $10^9$ ) to  $20^{20}$ . Thus, with libraries of  $10^7$  to  $10^8$  different molecules the present methods allow a "working" subset of a theoretically complete interaction library for 7 amino acids, and a subset of shapes for the  $20^{20}$  library. Thus, in a preferred embodiment, at least  $10^6$ , preferably at least  $10^7$ , more preferably at least  $10^8$  and most preferably at least  $10^9$  different sequences are simultaneously analyzed in the subject methods. Preferred methods maximize library size and diversity.

The candidate bioactive agents are combined or added to a cell or population of cells. Suitable cell types for different embodiments are outlined above. By "population of cells" herein is meant at least two cells, with at least about  $10^5$  being preferred, at least about  $10^6$  being particularly preferred, and at least about  $10^7$ ,  $10^8$  and  $10^9$  being especially preferred.

The candidate bioactive agent and the cells are combined. As will be appreciated by those in the art, this may accomplished in any number of ways, including adding the candidate agents to the surface of the cells, to the media containing the cells, or to a surface on which the cells are growing or in contact with; adding the agents into the cells, for example by using vectors that will introduce the agents into the cells (i.e. when the agents are nucleic acids or proteins).

In a preferred embodiment, the candidate bioactive agents are either nucleic acids or proteins (proteins in this context includes proteins, oligopeptides, and peptides) that are introduced into the host cells using retroviral vectors, as is generally outlined in PCT US97/01019 and PCT US97/01048, both of which are expressly incorporated by reference. Generally, a library of retroviral vectors is made using retroviral packaging cell lines that are helper-defective and are capable of producing all the necessary trans proteins, including gag, pol and env, and RNA molecules that have in cis the  $\Psi$  packaging signal. Briefly, the library is generated in a retrovirus DNA construct backbone; standard oligonucleotide synthesis is done to generate either the candidate agent or nucleic acid encoding a protein, for example a random peptide,



using techniques well known in the art. After generation of the DNA library, the library is cloned into a first primer. The first primer serves as a “cassette”, which is inserted into the retroviral construct. The first primer generally contains a number of elements, including for example, the required regulatory sequences (e.g. translation, transcription, promoters, etc), fusion partners, restriction endonuclease (cloning and subcloning) sites, stop codons (preferably in all three frames), regions of complementarity for second strand priming (preferably at the end of the stop codon region as minor deletions or insertions may occur in the random region), etc.

A second primer is then added, which generally consists of some or all of the complementarity region to prime the first primer and optional necessary sequences for a second unique restriction site for subcloning. DNA polymerase is added to make double-stranded oligonucleotides. The double-stranded oligonucleotides are cleaved with the appropriate subcloning restriction endonucleases and subcloned into the target retroviral vectors, described below.

Any number of suitable retroviral vectors may be used. Generally, the retroviral vectors may include: selectable marker genes under the control of internal ribosome entry sites (IRES) that greatly facilitates the selection of cells expressing peptides at uniformly high levels; and promoters driving expression of a second gene, placed in sense or anti-sense relative to the 5' LTR. Suitable selection genes include, but are not limited to, neomycin, blastocidin, bleomycin, puromycin, and hygromycin resistance genes, as well as self-fluorescent markers such as green fluorescent protein, enzymatic markers such as lacZ, and surface proteins such as CD8, etc.

Preferred vectors include a vector based on the murine stem cell virus (MSCV) (see Hawley et al., Gene Therapy 1:136 (1994)) and a modified MFG virus (Riviere et al., Genetics 92:6733 (1995)), and pBABE, outlined in the examples.

The retroviruses may include inducible and constitutive promoters for the expression of the candidate agent (to be distinguished from the IL-4 inducible  $\epsilon$  promoter). For example, there are situations wherein it is necessary to induce peptide expression only during certain phases of the selection process. A large number of both inducible and constitutive promoters are known.

In addition, it is possible to configure a retroviral vector to allow inducible expression of retroviral inserts after integration of a single vector in target cells; importantly, the entire system is contained within the single retrovirus. Tet-inducible retroviruses have been designed incorporating the Self-Inactivating (SIN) feature of 3' LTR enhancer/promoter retroviral deletion mutant (Hoffman et al., PNAS USA 93:5185 (1996)). Expression of this vector in cells is virtually undetectable in the presence of tetracycline or other active analogs. However, in the absence of Tet, expression is turned on to maximum within 48 hours after induction, with uniform increased expression of the whole population of cells that harbor the inducible retrovirus, indicating that expression is regulated uniformly within the infected cell population. A similar, related system uses a mutated Tet DNA-binding domain such that it bound DNA in the presence of Tet, and was removed in the absence of Tet. Either of these systems is suitable.

In a preferred embodiment, the candidate bioactive agents are linked to a fusion partner. By "fusion partner" or "functional group" herein is meant a sequence that is associated with the candidate bioactive agent, that confers upon all members of the library in that class a common function or ability. Fusion partners can be heterologous (i.e. not native to the host cell), or synthetic (not native to any cell). Suitable fusion partners include, but are not limited to: a) presentation structures, as defined below, which provide the candidate bioactive agents in a conformationally restricted or stable form; b) targeting sequences, defined below, which allow the localization of the candidate bioactive agent into a subcellular or extracellular compartment, particularly a nuclear localization sequence (NLS); c) rescue sequences as defined below, which allow the purification or isolation of either the candidate bioactive agents or the nucleic acids encoding them; d) stability sequences, which confer stability or

protection from degradation to the candidate bioactive agent or the nucleic acid encoding it, for example resistance to proteolytic degradation; e) dimerization sequences, to allow for peptide dimerization; f) reporter genes (preferably a labeling gene or a survival gene); or g) any combination of a), b), c), d), e), or f) as well as linker sequences as needed.

5

In a preferred embodiment, the fusion partner is a presentation structure. By "presentation structure" or grammatical equivalents herein is meant a sequence, which, when fused to candidate bioactive agents, causes the candidate agents to assume a conformationally restricted form. Proteins interact with each other largely through conformationally constrained domains. Although small peptides with freely rotating amino and carboxyl termini can have potent functions as is known in the art, the conversion of such peptide structures into pharmacologic agents is difficult due to the inability to predict side-chain positions for peptidomimetic synthesis. Therefore the presentation of peptides in conformationally constrained structures will benefit both the later generation of pharmaceuticals and will also likely lead to higher affinity interactions of the peptide with the target protein. This fact has been recognized in the combinatorial library generation systems using biologically generated short peptides in bacterial phage systems. A number of workers have constructed small domain molecules in which one might present randomized peptide structures.

10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95  
100  
105  
110  
115  
120  
125  
130  
135  
140  
145  
150  
155  
160  
165  
170  
175  
180  
185  
190  
195  
200  
205  
210  
215  
220  
225  
230  
235  
240  
245  
250  
255  
260  
265  
270  
275  
280  
285  
290  
295  
300  
305  
310  
315  
320  
325  
330  
335  
340  
345  
350  
355  
360  
365  
370  
375  
380  
385  
390  
395  
400  
405  
410  
415  
420  
425  
430  
435  
440  
445  
450  
455  
460  
465  
470  
475  
480  
485  
490  
495  
500

20

While the candidate bioactive agents may be either nucleic acid or peptides, presentation structures are preferably used with peptide candidate agents. Thus, synthetic presentation structures, i.e. artificial polypeptides, are capable of presenting a randomized peptide as a conformationally-restricted domain. Generally such presentation structures comprise a first portion joined to the N-terminal end of the randomized peptide, and a second portion joined to the C-terminal end of the peptide; that is, the peptide is inserted into the presentation structure, although variations may be made, as outlined below. To increase the functional isolation of the randomized expression product, the presentation structures are selected or designed to have minimal biological activity when expressed in the target cell.

25

30

Preferred presentation structures maximize accessibility to the peptide by presenting it on an exterior loop. Accordingly, suitable presentation structures include, but are not limited to, minibody structures, loops on beta-sheet turns and coiled-coil stem structures in which residues not critical to structure are randomized, zinc-finger domains, cysteine-linked (disulfide) structures, transglutaminase linked structures, cyclic peptides, B-loop structures, helical barrels or bundles, leucine zipper motifs, etc.

In a preferred embodiment, the presentation structure is a coiled-coil structure, allowing the presentation of the randomized peptide on an exterior loop. See, for example, Myszkowski et al., Biochem. 33:2362-2373 (1994), hereby incorporated by reference). Using this system investigators have isolated peptides capable of high affinity interaction with the appropriate target. In general, coiled-coil structures allow for between 6 to 20 randomized positions.

A preferred coiled-coil presentation structure is as follows:

MGCAALESEVSALESEVASLESEVAAL**GRGDMP**LAAVKSKLSAVKSKLASVKSKLA  
ACGPP. The underlined regions represent a coiled-coil leucine zipper region defined previously (see Martin et al., EMBO J. 13(22):5303-5309 (1994), incorporated by reference). The bolded GRGDMP region represents the loop structure and when appropriately replaced with randomized peptides (i.e. candidate bioactive agents, generally depicted herein as (X)<sub>n</sub>, where X is an amino acid residue and n is an integer of at least 5 or 6) can be of variable length. The replacement of the bolded region is facilitated by encoding restriction endonuclease sites in the underlined regions, which allows the direct incorporation of randomized oligonucleotides at these positions. For example, a preferred embodiment generates a XhoI site at the double underlined LE site and a HindIII site at the double-underlined KL site.

In a preferred embodiment, the presentation structure is a minibody structure. A "minibody" is essentially composed of a minimal antibody complementarity region. The minibody presentation structure generally provides two randomizing regions that in the folded protein are presented along a single face of the tertiary structure. See for example Bianchi et al., J.

Mol. Biol. 236(2):649-59 (1994), and references cited therein, all of which are incorporated by reference). Investigators have shown this minimal domain is stable in solution and have used phage selection systems in combinatorial libraries to select minibodies with peptide regions exhibiting high affinity,  $K_d = 10^{-7}$ , for the pro-inflammatory cytokine IL-6.

5

A preferred minibody presentation structure is as follows:

MGRNSQATS**GFT**F**SHFY**MEWVRGGEYIAASR**HKHNKY**TTEYSASVKGRYIVSRDT  
SQSILYLQKKKGPP. The bold, underline regions are the regions which may be randomized.

The italicized phenylalanine must be invariant in the first randomizing region. The entire peptide is cloned in a three-oligonucleotide variation of the coiled-coil embodiment, thus allowing two different randomizing regions to be incorporated simultaneously. This embodiment utilizes non-palindromic BstXI sites on the termini.

In a preferred embodiment, the presentation structure is a sequence that contains generally two cysteine residues, such that a disulfide bond may be formed, resulting in a conformationally constrained sequence. This embodiment is particularly preferred when secretory targeting sequences are used. As will be appreciated by those in the art, any number of random sequences, with or without spacer or linking sequences, may be flanked with cysteine residues. In other embodiments, effective presentation structures may be generated by the random regions themselves. For example, the random regions may be "doped" with cysteine residues which, under the appropriate redox conditions, may result in highly crosslinked structured conformations, similar to a presentation structure. Similarly, the randomization regions may be controlled to contain a certain number of residues to confer  $\beta$ -sheet or  $\alpha$ -helical structures.

25

In a preferred embodiment, the fusion partner is a targeting sequence that targets the candidate bioactive agent to a particular subcellular location. As will be appreciated by those in the art, the localization of proteins within a cell is a simple method for increasing effective concentration and determining function. The concentration of a protein can also be simply increased by nature of the localization. Shuttling the proteins into the nucleus confines them

30

to a smaller space thereby increasing concentration. While other targeting sequences such as targeting sequences to the Golgi, endoplasmic reticulum, nuclear membrane, mitochondria, secretory vesicles, lysosome, and cellular membrane may be used, a preferred embodiment uses targeting sequences to the nucleus, i.e. a nuclear localization signal (NLS).

5

In a preferred embodiment, the targeting sequence is a nuclear localization signal (NLS). NLSs are generally short, positively charged (basic) domains that serve to direct the entire protein in which they occur to the cell's nucleus. Numerous NLS amino acid sequences have been reported including single basic NLS's such as that of the SV40 (monkey virus) large T Antigen (Pro Lys Lys Lys Arg Lys Val), Kalderon (1984), et al., Cell, 39:499-509; the human retinoic acid receptor- $\beta$  nuclear localization signal (ARRRRP); NF $\kappa$ B p50 (EEVQRKRQKL; Ghosh et al., Cell 62:1019 (1990); NF $\kappa$ B p65 (EEKRKRTYE; Nolan et al., Cell 64:961 (1991); and others (see for example Bouliskas, J. Cell. Biochem. 55(1):32-58 (1994), hereby incorporated by reference) and double basic NLS's exemplified by that of the Xenopus (African clawed toad) protein, nucleoplasmin (Ala Val Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys Leu Asp), Dingwall, et al., Cell, 30:449-458, 1982 and Dingwall, et al., J. Cell Biol., 107:641-849; 1988). Numerous localization studies have demonstrated that NLSs incorporated in synthetic peptides or grafted onto reporter proteins not normally targeted to the cell nucleus cause these peptides and reporter proteins to be concentrated in the nucleus. See, for example, Dingwall, and Laskey, Ann. Rev. Cell Biol., 2:367-390, 1986; Bonnerot, et al., Proc. Natl. Acad. Sci. USA, 84:6795-6799, 1987; Galileo, et al., Proc. Natl. Acad. Sci. USA, 87:458-462, 1990.

10  
15  
20

25

In a preferred embodiment, the fusion partner is a rescue sequence. A rescue sequence is a sequence which may be used to purify or isolate either the candidate agent or the nucleic acid encoding it. Thus, for example, peptide rescue sequences include purification sequences such as the His<sub>6</sub> tag for use with Ni affinity columns and epitope tags for detection, immunoprecipitation or FACS (fluorescence-activated cell sorting). Suitable epitope tags include myc (for use with the commercially available 9E10 antibody), the BSP biotinylation target sequence of the bacterial enzyme BirA, flu tags, lacZ, and GST.

30

Alternatively, the rescue sequence may be a unique oligonucleotide sequence which serves as a probe target site to allow the quick and easy isolation of the retroviral construct, via PCR, related techniques, or hybridization.

5 In a preferred embodiment, the fusion partner is a stability sequence to confer stability to the candidate bioactive agent or the nucleic acid encoding it. Thus, for example, peptides may be stabilized by the incorporation of glycines after the initiation methionine (MG or MGG0), for protection of the peptide to ubiquitination as per Varshavsky's N-End Rule, thus conferring long half-life in the cytoplasm. Similarly, two prolines at the C-terminus impart peptides that are largely resistant to carboxypeptidase action. The presence of two glycines prior to the prolines impart both flexibility and prevent structure initiating events in the di-proline to be propagated into the candidate peptide structure. Thus, preferred stability sequences are as follows:  $MG(X)_nGGPP$ , where X is any amino acid and n is an integer of at least four.

10  
15 In one embodiment, the fusion partner is a dimerization sequence. A dimerization sequence allows the non-covalent association of one random peptide to another random peptide, with sufficient affinity to remain associated under normal physiological conditions. This effectively allows small libraries of random peptides (for example,  $10^4$ ) to become large libraries if two peptides per cell are generated which then dimerize, to form an effective library of  $10^8$  ( $10^4 \times 10^4$ ). It also allows the formation of longer random peptides, if needed, or more structurally complex random peptide molecules. The dimers may be homo- or heterodimers.

20  
25 Dimerization sequences may be a single sequence that self-aggregates, or two sequences, each of which is generated in a different retroviral construct. That is, nucleic acids encoding both a first random peptide with dimerization sequence 1, and a second random peptide with dimerization sequence 2, such that upon introduction into a cell and expression of the nucleic acid, dimerization sequence 1 associates with dimerization sequence 2 to form a new random peptide structure.

5

20

25

30

In a preferred embodiment, the fusion partner includes a linker or tethering sequence, as generally described in PCT US 97/01019, that can allow the candidate agents to interact with potential targets unhindered. For example, when the candidate bioactive agent is a peptide,



useful linkers include glycine-serine polymers (including, for example, (GS)<sub>n</sub>, (GSGGS)<sub>n</sub> and (GGGS)<sub>n</sub>, where n is an integer of at least one), glycine-alanine polymers, alanine-serine polymers, and other flexible linkers such as the tether for the shaker potassium channel, and a large variety of other flexible linkers, as will be appreciated by those in the art. Glycine-serine polymers are preferred since both of these amino acids are relatively unstructured, and therefore may be able to serve as a neutral tether between components. Secondly, serine is hydrophilic and therefore able to solubilize what could be a globular glycine chain. Third, similar chains have been shown to be effective in joining subunits of recombinant proteins such as single chain antibodies.

In addition, the fusion partners, including presentation structures, may be modified, randomized, and/or matured to alter the presentation orientation of the randomized expression product. For example, determinants at the base of the loop may be modified to slightly modify the internal loop peptide tertiary structure, which maintaining the randomized amino acid sequence.

In a preferred embodiment, combinations of fusion partners are used. Thus, for example, any number of combinations of presentation structures, targeting sequences, rescue sequences, and stability sequences may be used, with or without linker sequences.

Thus, candidate agents can include these components, and may then be used to generate a library of fragments, each containing a different random nucleotide sequence that may encode a different peptide. The ligation products are then transformed into bacteria, such as *E. coli*, and DNA is prepared from the resulting library, as is generally outlined in Kitamura, PNAS USA 92:9146-9150 (1995), hereby expressly incorporated by reference.

Delivery of the library DNA into a retroviral packaging system results in conversion to infectious virus. Suitable retroviral packaging system cell lines include, but are not limited to, the Bing and BOSC23 cell lines described in WO 94/19478; Soneoka et al., Nucleic Acid Res. 23(4):628 (1995); Finer et al., Blood 83:43 (1994); Pheonix packaging lines such as

PhiNX-eco and PhiNX-ampho, described below; 292T + gag-pol and retrovirus envelope; PA317; and cell lines outlined in Markowitz et al., Virology 167:400 (1988), Markowitz et al., J. Virol. 62:1120 (1988), Li et al., PNAS USA 93:11658 (1996), Kinsella et al., Human Gene Therapy 7:1405 (1996), all of which are incorporated by reference. Preferred systems include PhiNX-eco and PhiNX-ampho or similar cell lines, disclosed in PCT US97/01019.

In general, the candidate agents are added to the cells under reaction conditions that favor agent-target interactions. Generally, this will be physiological conditions. Incubations may be performed at any temperature which facilitates optimal activity, typically between 4 and 40°C. Incubation periods are selected for optimum activity, but may also be optimized to facilitate rapid high through put screening. Typically between 0.1 and 1 hour will be sufficient. Excess reagent is generally removed or washed away.

A variety of other reagents may be included in the assays. These include reagents like salts, neutral proteins, e.g. albumin, detergents, etc which may be used to facilitate optimal protein-protein binding and/or reduce non-specific or background interactions. Also reagents that otherwise improve the efficiency of the assay, such as protease inhibitors, nuclease inhibitors, anti-microbial agents, etc., may be used. The mixture of components may be added in any order that provides for the requisite binding.

Once the candidate agents have been introduced or combined with the cells containing the fusion constructs, the IL-4 inducible  $\epsilon$  promoter is induced. Alternatively, the promoter is induced prior to the addition of the candidate bioactive agents, or simultaneously. This is generally done as is known in the art, and involves the addition of IL-4 or IL-13 to the cells at a concentration of not less than 5 units/ml with 200 units/ml being most preferred. Addition of IL-4 or IL-13 is usually 24-48 hours after the bioactive agents are added.

The presence or absence of the reporter gene is then detected. This may be done in a number of ways, as will be appreciated by those in the art, and will depend in part on the reporter gene. For example, cells expressing a label reporter gene, such as GFP, can be distinguished

from those not expressing the gene, and preferably sorted (enriched by FACS) on this basis. Similarly, cells expressing the death gene will die, leaving only cells that have inhibited promotion of the expression of the gene, etc. In general, the cells that express the reporter gene (i.e. non-inhibited IL-4 inducible  $\epsilon$  promoter) and separated from those that do not (i.e. the IL-4 inducible  $\epsilon$  promoter was inhibited). This may be done using FACS, lysis selection using complements, cell cloning, scanning by a Fluorimager, growth under drug resistance, enhanced growth, etc.

In a preferred embodiment, for example when the reporter gene is a death gene, sorting of cells containing bioactive agents that inhibit the IL-4 inducible  $\epsilon$  promoter (and thus do not turn on the death gene) from those cells that contain candidate agents that do not inhibit the promoter is simple: only those surviving cells contain such an agent.

In a preferred embodiment, the presence or absence of the reporter gene is determined using a fluorescent-activated cell sorter (FACS). In general, the expression of the reporter gene comprising a label (or allowing the use of a label) is optimized to allow for efficient enrichment by FACS. Thus, for example, in general, 10 to 1000 fluores per sorting event are needed; i.e. per cell, with from about 100 to 1000 being preferred, and from 500 to 1000 being especially preferred. This can be accomplished by amplifying the signal per reporter gene, i.e. have each second label comprise multiple fluores, or by having a high density of reporter genes per cell; or a combination of both.

In a preferred embodiment, the cells are sorted at very high speeds, for example greater than about 5,000 sorting events per sec, with greater than about 10,000 sorting events per sec being preferred, and greater than about 25,000 sorting events per second being particularly preferred, with speeds of greater than about 50,000 to 100,000 being especially preferred. The use of multiple laser paths allows sort accuracy of 1 in  $10^6$  with better than 70% accuracy.

The sorting results in a population of cells containing the reporter protein (i.e. the promoter was not inhibited) and at least one population of cells without the reporter protein (i.e. the promoter was inhibited). The absence of the reporter protein is indicative that at least one candidate bioactive agent is a bioactive agent that inhibits the IL-4 inducible  $\epsilon$  promoter.

In addition to screening methods utilizing the reporter constructs described above, the invention also provides methods for screening candidate agents for the ability to modulate IgE production. By “modulating IgE production” herein is meant either an increase or a decrease in IgE production, as quantified by the amount of IgE protein made. In this embodiment, cells that have already switched to the  $\epsilon$  heavy chain region can no longer be blocked at the earlier phase of IgE production. This is especially important for memory B cells that maintain their capacity to secrete IgE and are long lived. Thus, in this embodiment, candidate agents are screened to identify compounds that can block IgE at the level of  $\epsilon$  heavy chain transcription, translation, assembly and trafficking, to prevent the terminal stages of IgE production. In this embodiment, a candidate bioactive agent is combined with a cell capable of expressing IgE, preferably surface IgE. Preferred cells include, but are not limited to, cells that produce surface IgE such as the U266 cell line (Lagging, et al., “Distribution of Plasma Cell Markers and Intracellular IgE in Cell Line U266,” *Immunology Letters* 49:71 (1996)).

The candidate agent and the cells are combined, as outlined above, and the cells screened for alterations in the amount of IgE produced, as compared to the amount produced in the absence of the candidate bioactive agent. This may be done using standard IgE labeling techniques, including, but not limited to, the use of anti-IgE antibodies, that may be either directly or indirectly labeled, for example through the use of fluorescent anti-IgE antibodies or fluorescent secondary antibodies, and through the use of IgE fusion proteins, as outlined below.

In a preferred embodiment, the amount of IgE produced is determined through the use of IgE fusion proteins; that is, the IgE is produced as a fusion protein comprising the IgE protein,

specifically at least the  $\epsilon$  heavy chain, and a detectable protein such as is generally outlined above for label reporter genes. In a preferred embodiment, gene "knock in" cell lines are produced, as outlined above and shown in the Figures. In this embodiment, a first label gene, such as the gene for green fluorescent protein (GFP), is fused to the secretory exon of IgE to label secretory IgE heavy chains green. In a preferred embodiment, a second label gene, such as the gene for blue fluorescent protein (BFP), is attached to the M2 exon to label membrane IgE heavy chains blue. This is preferred as it allows discrimination between mRNA processing and translation of secretory versus membrane  $\epsilon$ -heavy chain transcripts. Suitable label genes for this embodiment include, but are not limited to, GFP, BFP, YFP and RFP.

Accordingly, the present invention provides cell lines that produce fusion proteins comprising IgE (either secreted or membrane bound) fused to a label protein, preferably a fluorescent protein.

In yet another preferred embodiment, the invention provides methods of identifying proteins that bind to all or part of the switch  $\epsilon$  region (Figure 2B). The general idea is to use a "one hybrid" system to identify proteins that bind to all or part of the switch  $\epsilon$  region. To this end, the present invention provides compositions comprising a test vector and a reporter vector, and cells containing these vectors. These cells may be yeast, such as YM4271 or any yeast cell lines that reporter constructs can be inserted into.

By "vector" or "episome" herein is meant a replicon used for the transformation of host cells. The vectors may be either self-replicating extrachromosomal vectors ("plasmids") or vectors which integrate into a host genome. A preferred embodiment utilizes retroviral vectors, as is more fully described below.

Suitable vectors will depend on the host cells used. For use of the system in yeast, suitable vectors are known in the art and include, but are not limited to, pHisi-1 and pLacZi (Clontech Cat #K1603-1) (Li, et al., "Isolation of ORC6, A Component of the Yeast Origin of Recognition Complex By a One-Hybrid System," *Science* 262:1870-1873 (1993); Liu, et

al. "Identifying DNA-Binding Sites and Analyzing DNA-Binding Domains Using a Yeast Selection System," In: *Methods: A Companion to Methods in Enzymology* 5:125-137 (1993), Luo, et al., "Cloning and Analysis of DNA-Binding Proteins By Yeast One-Hybrid and One-Two-Hybrid Systems," *Biotechniques* 20:564-568 (1996), and Strubin, et al., "OBF-1, A Novel B Cell-Specific Coactivator That Stimulates Immunoglobulin Promoter Activity Through Association with Octamer-Binding Proteins," *Cell* 80:497-506 (1995)). Yeast expression systems are well known in the art, and include expression vectors for *Saccharomyces cerevisiae*, *Candida albicans* and *C. maltosa*, *Hansenula polymorpha*, *Kluyveromyces fragilis* and *K. lactis*, *Pichia guillerimondii* and *P. pastoris*, *Schizosaccharomyces pombe*, and *Yarrowia lipolytica*. Preferred promoter sequences for expression in yeast include the inducible GAL1,10 promoter, the promoters from alcohol dehydrogenase, enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase, hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, pyruvate kinase, and the acid phosphatase gene. Yeast selectable markers include ADE2, HIS4, LEU2, TRP1, and ALG7, which confers resistance to tunicamycin; the neomycin phosphotransferase gene, which confers resistance to G418; and the CUP1 gene, which allows yeast to grow in the presence of copper ions.

For non-retroviral mammalian cell embodiments, suitable vectors are derived from any number of known vectors, including, but not limited to, pCEP4 (Invitrogen), pCI-NEO (Promega), and pBI-EGFP (Clontech). Basically, any mammalian expression vectors with strong promoters such as CMV can be used to construct test vectors.

In a preferred embodiment, one or more retroviral vectors are used. Currently, the most efficient gene transfer methodologies harness the capacity of engineered viruses, such as retroviruses, to bypass natural cellular barriers to exogenous nucleic acid uptake. The use of recombinant retroviruses was pioneered by Richard Mulligan and David Baltimore with the Psi-2 lines and analogous retrovirus packaging systems, based on NIH 3T3 cells (see Mann et al., *Cell* 33:153-159 (1993), hereby incorporated by reference). Such helper-defective packaging lines are capable of producing all the necessary trans proteins -gag, pol, and env-

that are required for packaging, processing, reverse transcription, and integration of recombinant genomes. Those RNA molecules that have in cis the  $\psi$  packaging signal are packaged into maturing virions.

5 Retroviruses are preferred for a number of reasons. First, their derivation is easy. Second, unlike Adenovirus-mediated gene delivery, expression from retroviruses is long-term (adenoviruses do not integrate). Adeno-associated viruses have limited space for genes and regulatory units and there is some controversy as to their ability to integrate. Retroviruses therefore offer the best current compromise in terms of long-term expression, genomic flexibility, and stable integration, among other features. The main advantage of retroviruses is that their integration into the host genome allows for their stable transmission through cell division. This ensures that in cell types which undergo multiple independent maturation steps, such as hematopoietic cell progression, the retrovirus construct will remain resident and continue to express. In addition, transfection efficiencies can be extremely high, thus obviating the need for selection genes in some cases.

10  
15  
20 A particularly well suited retroviral transfection system is described in Mann et al., supra: Pear et al., PNAS USA 90(18):8392-6 (1993); Kitamura et al., PNAS USA 92:9146-9150 (1995); Kinsella et al., Human Gene Therapy 7:1405-1413; Hofmann et al., PNAS USA 93:5185-5190; Choate et al., Human Gene Therapy 7:2247 (1996); WO 94/19478; PCT US97/01019, and references cited therein, all of which are incorporated by reference.

25 Any number of suitable retroviral vectors may be used. Preferred retroviral vectors include a vector based on the murine stem cell virus (MSCV) (see Hawley et al., Gene Therapy 1:136 (1994)) and a modified MFG virus (Riviere et al., Genetics 92:6733 (1995)), and pBABE (see PCT US97/01019, incorporated by reference). Particularly preferred vectors are shown in Figure 11.

30 As for the other vectors, the retroviral vectors may include inducible and constitutive promoters. Constitutive promoters are preferred for the bait and test vectors, and include, but

are not limited to, CMV, SV40, Sr $\alpha$ , RSV, and TK. Similarly, the reporter vector promoter is associated with at least one copy of an operator, as outlined herein.

In addition, it is possible to configure a retroviral vector to allow expression of bait genes or test genes after integration of a bait or test vector in target cells. For example, Tet-inducible retroviruses can be used to express bait or test genes (Hoffman et al., PNAS USA 93:5185 (1996)). Expression of this vector in cells is virtually undetectable in the presence of tetracycline or other active analogs. However, in the absence of Tet, expression is turned on to maximum within 48 hours after induction, with uniform increased expression of the whole population of cells that harbor the inducible retrovirus, indicating that expression is regulated uniformly within the infected cell population. A similar, related system uses a mutated Tet DNA-binding domain such that it bound DNA in the presence of Tet, and was removed in the absence of Tet. Either of these systems is suitable.

Generally, these expression vectors include transcriptional and translational regulatory nucleic acid operably linked to nucleic acids which are to be expressed. "Operably linked" in this context means that the transcriptional and translational regulatory nucleic acid is positioned relative to any coding sequences in such a manner that transcription is initiated. Generally, this will mean that the promoter and transcriptional initiation or start sequences are positioned 5' to the coding region. The transcriptional and translational regulatory nucleic acid will generally be appropriate to the host cell used, as will be appreciated by those in the art. Numerous types of appropriate expression vectors, and suitable regulatory sequences, are known in the art for a variety of host cells.

In general, the transcriptional and translational regulatory sequences may include, but are not limited to, promoter sequences, ribosomal binding sites, transcriptional start and stop sequences, translational start and stop sequences, and enhancer or activator sequences. In a preferred embodiment, the regulatory sequences include a promoter and transcriptional start and stop sequences.



Promoter sequences encode either constitutive or inducible promoters. The promoters may be either naturally occurring promoters, hybrid or synthetic promoters. Hybrid promoters, which combine elements of more than one promoter, are also known in the art, and are useful in the present invention.

5

In general, the vectors of the present invention utilize two different types of promoters. In a preferred embodiment, the promoters on the bait and test vectors are constitutive, and drive the expression of the fusion proteins and selection genes, if applicable, at a high level. However, it is possible to utilize inducible promoters for the fusion constructs and selection genes, if necessary.

10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000  
1001  
1002  
1003  
1004  
1005  
1006  
1007  
1008  
1009  
1010  
1011  
1012  
1013  
1014  
1015  
1016  
1017  
1018  
1019  
1020  
1021  
1022  
1023  
1024  
1025  
1026  
1027  
1028  
1029  
1030  
1031  
1032  
1033  
1034  
1035  
1036  
1037  
1038  
1039  
1040  
1041  
1042  
1043  
1044  
1045  
1046  
1047  
1048  
1049  
1050  
1051  
1052  
1053  
1054  
1055  
1056  
1057  
1058  
1059  
1060  
1061  
1062  
1063  
1064  
1065  
1066  
1067  
1068  
1069  
1070  
1071  
1072  
1073  
1074  
1075  
1076  
1077  
1078  
1079  
1080  
1081  
1082  
1083  
1084  
1085  
1086  
1087  
1088  
1089  
1090  
1091  
1092  
1093  
1094  
1095  
1096  
1097  
1098  
1099  
1100  
1101  
1102  
1103  
1104  
1105  
1106  
1107  
1108  
1109  
1110  
1111  
1112  
1113  
1114  
1115  
1116  
1117  
1118  
1119  
1120  
1121  
1122  
1123  
1124  
1125  
1126  
1127  
1128  
1129  
1130  
1131  
1132  
1133  
1134  
1135  
1136  
1137  
1138  
1139  
1140  
1141  
1142  
1143  
1144  
1145  
1146  
1147  
1148  
1149  
1150  
1151  
1152  
1153  
1154  
1155  
1156  
1157  
1158  
1159  
1160  
1161  
1162  
1163  
1164  
1165  
1166  
1167  
1168  
1169  
1170  
1171  
1172  
1173  
1174  
1175  
1176  
1177  
1178  
1179  
1180  
1181  
1182  
1183  
1184  
1185  
1186  
1187  
1188  
1189  
1190  
1191  
1192  
1193  
1194  
1195  
1196  
1197  
1198  
1199  
1200  
1201  
1202  
1203  
1204  
1205  
1206  
1207  
1208  
1209  
1210  
1211  
1212  
1213  
1214  
1215  
1216  
1217  
1218  
1219  
1220  
1221  
1222  
1223  
1224  
1225  
1226  
1227  
1228  
1229  
1230  
1231  
1232  
1233  
1234  
1235  
1236  
1237  
1238  
1239  
1240  
1241  
1242  
1243  
1244  
1245  
1246  
1247  
1248  
1249  
1250  
1251  
1252  
1253  
1254  
1255  
1256  
1257  
1258  
1259  
1260  
1261  
1262  
1263  
1264  
1265  
1266  
1267  
1268  
1269  
1270  
1271  
1272  
1273  
1274  
1275  
1276  
1277  
1278  
1279  
1280  
1281  
1282  
1283  
1284  
1285  
1286  
1287  
1288  
1289  
1290  
1291  
1292  
1293  
1294  
1295  
1296  
1297  
1298  
1299  
1300  
1301  
1302  
1303  
1304  
1305  
1306  
1307  
1308  
1309  
1310  
1311  
1312  
1313  
1314  
1315  
1316  
1317  
1318  
1319  
1320  
1321  
1322  
1323  
1324  
1325  
1326  
1327  
1328  
1329  
1330  
1331  
1332  
1333  
1334  
1335  
1336  
1337  
1338  
1339  
1340  
1341  
1342  
1343  
1344  
1345  
1346  
1347  
1348  
1349  
1350  
1351  
1352  
1353  
1354  
1355  
1356  
1357  
1358  
1359  
1360  
1361  
1362  
1363  
1364  
1365  
1366  
1367  
1368  
1369  
1370  
1371  
1372  
1373  
1374  
1375  
1376  
1377  
1378  
1379  
1380  
1381  
1382  
1383  
1384  
1385  
1386  
1387  
1388  
1389  
1390  
1391  
1392  
1393  
1394  
1395  
1396  
1397  
1398  
1399  
1400  
1401  
1402  
1403  
1404  
1405  
1406  
1407  
1408  
1409  
1410  
1411  
1412  
1413  
1414  
1415  
1416  
1417  
1418  
1419  
1420  
1421  
1422  
1423  
1424  
1425  
1426  
1427  
1428  
1429  
1430  
1431  
1432  
1433  
1434  
1435  
1436  
1437  
1438  
1439  
1440  
1441  
1442  
1443  
1444  
1445  
1446  
1447  
1448  
1449  
1450  
1451  
1452  
1453  
1454  
1455  
1456  
1457  
1458  
1459  
1460  
1461  
1462  
1463  
1464  
1465  
1466  
1467  
1468  
1469  
1470  
1471  
1472  
1473  
1474  
1475  
1476  
1477  
1478  
1479  
1480  
1481  
1482  
1483  
1484  
1485  
1486  
1487  
1488  
1489  
1490  
1491  
1492  
1493  
1494  
1495  
1496  
1497  
1498  
1499  
1500  
1501  
1502  
1503  
1504  
1505  
1506  
1507  
1508  
1509  
1510  
1511  
1512  
1513  
1514  
1515  
1516  
1517  
1518  
1519  
1520  
1521  
1522  
1523  
1524  
1525  
1526  
1527  
1528  
1529  
1530  
1531  
1532  
1533  
1534  
1535  
1536  
1537  
1538  
1539  
1540  
1541  
1542  
1543  
1544  
1545  
1546  
1547  
1548  
1549  
1550  
1551  
1552  
1553  
1554  
1555  
1556  
1557  
1558  
1559  
1560  
1561  
1562  
1563  
1564  
1565  
1566  
1567  
1568  
1569  
1570  
1571  
1572  
1573  
1574  
1575  
1576  
1577  
1578  
1579  
1580  
1581  
1582  
1583  
1584  
1585  
1586  
1587  
1588  
1589  
1590  
1591  
1592  
1593  
1594  
1595  
1596  
1597  
1598  
1599  
1600  
1601  
1602  
1603  
1604  
1605  
1606  
1607  
1608  
1609  
1610  
1611  
1612  
1613  
1614  
1615  
1616  
1617  
1618  
1619  
1620  
1621  
1622  
1623  
1624  
1625  
1626  
1627  
1628  
1629  
1630  
1631  
1632  
1633  
1634  
1635  
1636  
1637  
1638  
1639  
1640  
1641  
1642  
1643  
1644  
1645  
1646  
1647  
1648  
1649  
1650  
1651  
1652  
1653  
1654  
1655  
1656  
1657  
1658  
1659  
1660  
1661  
1662  
1663  
1664  
1665  
1666  
1667  
1668  
1669  
1670  
1671  
1672  
1673  
1674  
1675  
1676  
1677  
1678  
1679  
1680  
1681  
1682  
1683  
1684  
1685  
1686  
1687  
1688  
1689  
1690  
1691  
1692  
1693  
1694  
1695  
1696  
1697  
1698  
1699  
1700  
1701  
1702  
1703  
1704  
1705  
1706  
1707  
1708  
1709  
1710  
1711  
1712  
1713  
1714  
1715  
1716  
1717  
1718  
1719  
1720  
1721  
1722  
1723  
1724  
1725  
1726  
1727  
1728  
1729  
1730  
1731  
1732  
1733  
1734  
1735  
1736  
1737  
1738  
1739  
1740  
1741  
1742  
1743  
1744  
1745  
1746  
1747  
1748  
1749  
1750  
1751  
1752  
1753  
1754  
1755  
1756  
1757  
1758  
1759  
1760  
1761  
1762  
1763  
1764  
1765  
1766  
1767  
1768  
1769  
1770  
1771  
1772  
1773  
1774  
1775  
1776  
1777  
1778  
1779  
1780  
1781  
1782  
1783  
1784  
1785  
1786  
1787  
1788  
1789  
1790  
1791  
1792  
1793  
1794  
1795  
1796  
1797  
1798  
1799  
1800  
1801  
1802  
1803  
1804  
1805  
1806  
1807  
1808  
1809  
1810  
1811  
1812  
1813  
1814  
1815  
1816  
1817  
1818  
1819  
1820  
1821  
1822  
1823  
1824  
1825  
1826  
1827  
1828  
1829  
1830  
1831  
1832  
1833  
1834  
1835  
1836  
1837  
1838  
1839  
1840  
1841  
1842  
1843  
1844  
1845  
1846  
1847  
1848  
1849  
1850  
1851  
1852  
1853  
1854  
1855  
1856  
1857  
1858  
1859  
1860  
1861  
1862  
1863  
1864  
1865  
1866  
1867  
1868  
1869  
1870  
1871  
1872  
1873  
1874  
1875  
1876  
1877  
1878  
1879  
1880  
1881  
1882  
1883  
1884  
1885  
1886  
1887  
1888  
1889  
1890  
1891  
1892  
1893  
1894  
1895  
1896  
1897  
1898  
1899  
1900  
1901  
1902  
1903  
1904  
1905  
1906  
1907  
1908  
1909  
1910  
1911  
1912  
1913  
1914  
1915  
1916  
1917  
1918  
1919  
1920  
1921  
1922  
1923  
1924  
1925  
1926  
1927  
1928  
1929  
1930  
1931  
1932  
1933  
1934  
1935  
1936  
1937  
1938  
1939  
1940  
1941  
1942  
1943  
1944  
1945  
1946  
1947  
1948  
1949  
1950  
1951  
1952  
1953  
1954  
1955  
1956  
1957  
1958  
1959  
1960  
1961  
1962  
1963  
1964  
1965  
1966  
1967  
1968  
1969  
1970  
1971  
1972  
1973  
1974  
1975  
1976  
1977  
1978  
1979  
1980  
1981  
1982  
1983  
1984  
1985  
1986  
1987  
1988  
1989  
1990  
1991  
1992  
1993  
1994  
1995  
1996  
1997  
1998  
1999  
2000  
2001  
2002  
2003  
2004  
2005  
2006  
2007  
2008  
2009  
2010  
2011  
2012  
2013  
2014  
2015  
2016  
2017  
2018  
2019  
2020  
2021  
2022  
2023  
2024  
2025  
2026  
2027  
2028  
2029  
2030  
2031  
2032  
2033  
2034  
2035  
2036  
2037  
2038  
2039  
2040  
2041  
2042  
2043  
2044  
2045  
2046  
2047  
2048  
2049  
2050  
2051  
2052  
2053  
2054  
2055  
2056  
2057  
2058  
2059  
2060  
2061  
2062  
2063  
2064  
2065  
2066  
2067  
2068  
2069  
2070  
2071  
2072  
2073  
2074  
2075  
2076  
2077  
2078  
2079  
2080  
2081  
2082  
2083  
2084  
2085  
2086  
2087  
2088  
2089  
2090  
2091  
2092  
2093  
2094  
2095  
2096  
2097  
2098  
2099  
2100  
2101  
2102  
2103  
2104  
2105  
2106  
2107  
2108  
2109  
2110  
2111  
2112  
2113  
2114  
2115  
2116  
2117  
2118  
2119  
2120  
2121  
2122  
2123  
2124  
2125  
2126  
2127  
2128  
2129  
2130  
2131  
2132  
2133  
2134  
2135  
2136  
2137  
2138  
2139  
2140  
2141  
2142  
2143  
2144  
2145  
2146  
2147  
2148  
2149  
2150  
2151  
2152  
2153  
2154  
2155  
2156  
2157  
2158  
2159  
2160  
2161  
2162  
2163  
2164  
2165  
2166  
2167  
2168  
2169  
2170  
2171  
2172  
2173  
2174  
2175  
2176  
2177  
2178  
2179  
2180  
2181  
2182  
2183  
2184  
2185  
2186  
2187  
2188  
2189  
2190  
2191  
2192  
2193  
2194  
2195  
2196  
2197  
2198  
2199  
2200  
2201  
2202  
2203  
2204  
2205  
2

herein may be DNA, for example when extrachromosomal plasmids are the vectors, or RNA, for example when retroviral vectors are used. Generally, the sequences are directly linked together without any linking sequences, although in some embodiments linkers such as restriction endonuclease cloning sites or linkers encoding flexible amino acids such as glycine and serine linkers such as are known in the art are used. In a preferred embodiment, the first fusion gene comprises a first sequence encoding a transcriptional activation domain. By “transcriptional activator domain” herein is meant a proteinaceous domain which is able to activate transcription.

Suitable transcription activator domains include, but are not limited to, transcriptional activator domains from GAL4 (amino acids 1-147; see Fields et al., Nature 340:245 (1989), and Gill et al., PNAS USA 87:2127 (1990)), GCN4 (from *S. cerevisiae*, Hope et al., Cell 46:885 (1986)), ARD1 (from *S. cerevisiae*, Thukral et al., Mol. Cell. Biol. 9:2360 (1989)), the human estrogen receptor (Kumar et al., Cell 51:941 (1987)), VP16 (Triezenberg et al., Genes Dev. 2(6):718-729 (1988)), and B42 (Gyuris et al, Cell 1993), and NF-kB p65, and derivatives thereof which are functionally similar.

The fusion nucleic acid also includes a test nucleic acid, encoding a test protein. By “test protein” herein is meant a candidate protein which is to be tested for interaction with a bait protein. Protein in this context means proteins, oligopeptides, and peptides, i.e. at least two amino acids attached. In a preferred embodiment, the test protein sequence is one of a library of test protein sequences; that is, a library of test proteins is tested for binding to one or more bait proteins. The test protein sequences can be derived from genomic DNA, cDNA or can be random sequences. Alternatively, specific classes of test proteins may be tested. The library of test proteins or sequences encoding test proteins are incorporated into a library of test vectors, each or most containing a different test protein sequence.

In a preferred embodiment, the test protein sequences are derived from genomic DNA sequences. Generally, as will be appreciated by those in the art, genomic digests are cloned

into test vectors. The genomic library may be a complete library, or it may be fractionated or enriched as will be appreciated by those in the art.

In a preferred embodiment, the test protein sequences are derived from cDNA libraries. A cDNA library from any number of different cells may be used, and cloned into test vectors. As above, the cDNA library may be a complete library, or it may be fractionated or enriched in a number of ways.

In a preferred embodiment, the test protein sequences are random sequences. Generally, these will be generated from chemically synthesized oligonucleotides. Generally, random test proteins range in size from about 2 amino acids to about 100 amino acids, with from about 10 to about 50 amino acids being preferred. Fully random or "biased" random proteins may be used; that is, some positions within the sequence are either held constant or are selected from a limited number of possibilities. For example, in a preferred embodiment, the nucleotides or amino acid residues are randomized within a defined class, for example, of hydrophobic amino acids, hydrophilic residues, sterically biased (either small or large) residues, towards the creation of cysteines, for cross-linking, prolines for SH-3 domains, serines, threonines, tyrosines or histidines for phosphorylation sites, etc., for zinc fingers, SH-2 domains, stem loop structures, or to purines, or to reduce the chance of creation of a stop codon, etc.

The compositions of the invention also include reporter vectors. Generally, the test and reporter vectors are distinct, although as will be appreciated by those in the art, one or two independent vectors may be used. The reporter vectors comprise a first detectable or reporter gene and all or part of the switch  $\epsilon$  sequence, which functions as an operator site. That is, upon binding of a test protein to the switch  $\epsilon$  sequence (i.e. a protein-nucleic acid interaction), the transcriptional activator domain of the fusion protein will activate transcription and cause expression of the selectable or detectable gene(s). Thus, in this embodiment, the test protein functions essentially as a candidate agent.

In a preferred embodiment, the compositions are introduced into host cells to screen for protein-nucleic acid interactions. By "introduced into" or grammatical equivalents herein is meant that the nucleic acids enter the cells in a manner suitable for subsequent expression of the nucleic acid. The method of introduction is largely dictated by the targeted cell type and the composition of the vector. Exemplary methods include  $\text{CaPO}_4$  precipitation, liposome fusion, lipofectin®, electroporation, viral infection, etc. The vectors may stably integrate into the genome of the host cell (for example, with retroviral introduction for mammalian cells, outlined herein), or may exist either transiently or stably in the cytoplasm (i.e. through the use of traditional plasmids, utilizing standard regulatory sequences, selection markers, etc.).

The vectors can be introduced simultaneously, or sequentially in any order. In a preferred embodiment, host cells containing the reporter construct are generated first, and preferably the reporter vector is integrated into the genome of the host cell, for example, using a retroviral reporter vector. Once the components of the system are in the host cell, the cell is subjected to conditions under which the selectable markers and fusion proteins are expressed. If a test protein has sufficient affinity to the switch  $\epsilon$  region to activate transcription, the detectable protein is produced, and cells containing these proteins will survive drug selection and can be detected as outlined above. The detectable protein will be produced at a measurably higher level than in the absence of a protein-nucleic acid interaction. Thus the determination of a protein-nucleic acid interaction is generally done on the basis of the presence or absence of the detectable gene(s).

In a preferred embodiment, once a cell with an altered phenotype is detected, the cell is isolated from the plurality which do not have altered phenotypes. This may be done in any number of ways, as is known in the art, and will in some instances depend on the assay or screen. Suitable isolation techniques include, but are not limited to, drug selection, FACS, lysis selection using complement, cell cloning, scanning by Fluorimager, expression of a "survival" protein, induced expression of a cell surface protein or other molecule that can be rendered fluorescent or taggable for physical isolation; expression of an enzyme that changes a non-fluorescent molecule to a fluorescent one; overgrowth against a background of no or

slow growth; death of cells and isolation of DNA or other cell vitality indicator dyes; changes in fluorescent characteristics, etc. The preferred isolation techniques are drug selection and FACS based on the expression of the detectable gene, with a preferred embodiment utilizing both simultaneously.

5

Once a cell with a protein-nucleic acid interaction is detected and isolated, it is generally desirable to identify the test protein. In a preferred embodiment, the test protein nucleic acid and/or the test protein is isolated from the positive cell. This may be done in a number of ways. In a preferred embodiment, primers complementary to DNA regions common to the vector, or to specific components of the library such as a rescue sequence, are used to “rescue” the unique test sequence. Alternatively, the test protein is isolated using a rescue sequence. Thus, for example, rescue sequences comprising epitope tags or purification sequences may be used to pull out the test protein, using immunoprecipitation or affinity columns. Alternatively, the test protein may be detected using mass spectroscopy.

10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95  
100  
105  
110  
115  
120  
125  
130  
135  
140  
145  
150  
155  
160  
165  
170  
175  
180  
185  
190  
195  
200  
205  
210  
215  
220  
225  
230  
235  
240  
245  
250  
255  
260  
265  
270  
275  
280  
285  
290  
295  
300  
305  
310  
315  
320  
325  
330  
335  
340  
345  
350  
355  
360  
365  
370  
375  
380  
385  
390  
395  
400  
405  
410  
415  
420  
425  
430  
435  
440  
445  
450  
455  
460  
465  
470  
475  
480  
485  
490  
495  
500

Once a bioactive agent is identified, a number of things may be done. In a preferred embodiment, the characterization of the bioactive agent is done. This will proceed as will be appreciated by those in the art, and generally includes an analysis of the structure, identity, binding affinity and function of the agent. Depending on the type of agent, this may proceed in a number of ways. In a preferred embodiment, for example when the candidate agents have been introduced intracellularly using nucleic acid constructs, the candidate nucleic acid and/or the bioactive agent is isolated from the cells. This may be done in a number of ways. In a preferred embodiment, primers complementary to DNA regions common to the retroviral constructs, or to specific components of the library such as a rescue sequence, defined above, are used to “rescue” the unique random sequence. Alternatively, the bioactive agent is isolated using a rescue sequence. Thus, for example, rescue sequences comprising epitope tags or purification sequences may be used to pull out the bioactive agent, using immunoprecipitation or affinity columns. Alternatively, the peptide may be detected using mass spectroscopy.

30

Once rescued, the sequence of the bioactive agent and/or bioactive nucleic acid is determined. Similarly, candidate agents from other chemical classes can be identified and characterized, for example through the use of mass spectroscopy. This information can then be used in a number of ways.

5

In a preferred embodiment, the bioactive agent is resynthesized and reintroduced into the target cells, to verify the effect. This may be done using retroviruses, or alternatively using fusions to the HIV-1 Tat protein, and analogs and related proteins, which allows very high uptake into target cells. See for example, Fawell et al., PNAS USA 91:664 (1994); Frankel et al., Cell 55:1189 (1988); Savion et al., J. Biol. Chem. 256:1149 (1981); Derossi et al., J. Biol. Chem. 269:10444 (1994); and Baldin et al., EMBO J. 9:1511 (1990), all of which are incorporated by reference. Other techniques known in the art may be used as well.

In a preferred embodiment, the sequence of a bioactive agent is used to generate more candidate bioactive agents. For example, the sequence of the bioactive agent may be the basis of a second round of (biased) randomization, to develop bioactive agents with increased or altered activities. Alternatively, the second round of randomization may change the affinity of the bioactive agent. Furthermore, it may be desirable to put the identified random region of the bioactive agent into other presentation structures, or to alter the sequence of the constant region of the presentation structure, to alter the conformation/shape of the bioactive agent. It may also be desirable to "walk" around a potential binding site, in a manner similar to the mutagenesis of a binding pocket, by keeping one end of the ligand region constant and randomizing the other end to shift the binding of the peptide around.

Once identified and the biological activity is confirmed, the bioactive agent may be formulated. The compounds having the desired pharmacological activity may be administered in a physiologically acceptable carrier to a host, as previously described. The agents may be administered in a variety of ways, orally, parenterally *e.g.*, subcutaneously, intraperitoneally, intravascularly, etc. Depending upon the manner of introduction, the

compounds may be formulated in a variety of ways. The concentration of therapeutically active compound in the formulation may vary from about 0.1-100 wt.%.

The pharmaceutical compositions can be prepared in various forms, such as granules, tablets, pills, suppositories, capsules, suspensions, salves, lotions and the like. Pharmaceutical grade organic or inorganic carriers and/or diluents suitable for oral and topical use can be used to make up compositions containing the therapeutically-active compounds. Diluents known to the art include aqueous media, vegetable and animal oils and fats. Stabilizing agents, wetting and emulsifying agents, salts for varying the osmotic pressure or buffers for securing an adequate pH value, and skin penetration enhancers can be used as auxiliary agents.

The following examples serve to more fully describe the manner of using the above-described invention, as well as to set forth the best modes contemplated for carrying out various aspects of the invention. It is understood that these examples in no way serve to limit the true scope of this invention, but rather are presented for illustrative purposes. All references cited herein are incorporated by reference in their entirety.

## EXAMPLES

### Example 1

#### Construction of $\epsilon$ germline GFP/BFP knock-in cell lines

Three different IgM<sup>+</sup>, EBV<sup>-</sup> human B cells lines (CA-46, MC116, DND39, Figure 4) that produce  $\epsilon$  germline transcripts in the presence of IL-4 will be transfected with a germline  $\epsilon$  GFP or BFP knock-in construct (Figures 5B and 5C) and induced with IL-4. The cells will then be sorted by FACS for the appropriate reporter expression, GFP or BFP. Background (i.e. random integration) should be low since the construct must integrate downstream of an IL-4 inducible region in order to be activated. Homologous recombination of the reporter construct will be confirmed in fluorescent clones by genomic PCR using primers located within and immediately flanking the construct. For double knockouts, both GFP and BFP constructs will be transfected and cells sorted for expression of both reporters.

It is possible that activation with IL-4 to identify homologous recombined clones will result in events that move beyond the first phase of  $\epsilon$  switching, thus making the clones unusable for a screen identifying blockers of this first step. For this case, we have designed a more traditional construct containing an SV40 promoter-driven neomycin resistance gene which is flanked by loxP sites and inserted in the intron between the first and second  $\epsilon$  constant coding exons (Figure 5D). In addition, attached at the 3' end of the long arm is a BFP reporter gene driven by a constitutive promoter. B cell clones transfected with this construct will be selected for integration by culturing them in the presence of G418. The surviving cells lacking BFP will be sorted by FACS (the BFP at the 3' end will be preferentially deleted during the homologous recombination event). The remaining clones will be assessed for homologous recombination by PCR. Clones containing homologous recombined constructs will be exposed to the *cre* recombinase protein to mediate excision of the SV40 promoter/neomycin resistance gene in order to eliminate promoter interference and potential  $\epsilon$  promoter shutdown. Excision of the SV40 promoter/neomycin resistance gene fragment will be verified by subdividing clones into parent and daughter pools and re-selecting the latter pool in G418. The parental cells corresponding to G418 sensitive daughter cells will be subdivided again and tested for IL-4 inducible GFP expression. Parental stocks of the most inducible clones will be used for subsequent peptide screening. Production of the knock-in cell line using this approach would provide a continuous source of IL-4 inducible cells and would circumvent any down-regulation associated with IL-4 pre-treatment.

## Example 2

### Creation and screening of candidate bioactive agents in knock-in cell lines

A candidate bioactive agent library, in this case a peptide library, will be packaged into infectious viral particles as outlined below. A preferred library is a mixture of random peptide sequences with and without a nuclear localization sequence (NLS) upstream of a reporter gene to identify infected cells and relative peptide expression (see Figure 6).



Each screen will start with production of the primary retrovirus peptide library, as is generally shown in Figure 7. This is generally done as outlined in PCT US97/01019 and PCT US97/01048, both of which are expressly incorporated by reference. In general, this is done as follows. On day 1, the Phoenix cells are seeded in 10 cm plates at  $5 \times 10^6$  cells in 6 ml (DMEM + 10% FBS + Pen/Strep) per plate the day before transfection. Day 2: allow all reagents to reach room temperature 30 min. before starting. Add 50 mM chloroquine at 8  $\mu$ l/plate (50  $\mu$ M final) before preparing the transfection solution. Mix  $\text{CaPO}_4$  reagents in 15ml polypropylene tube: per plate: 10  $\mu$ g DNA, 122  $\mu$ l 2M  $\text{CaCl}_2$ , 876  $\mu$ l  $\text{H}_2\text{O}$ , 1.0ml 2X HBS. Add 2X HBS and depress the expulsion button completely to bubble air through the mix for 10 secs. Immediately add mixture gently dropwise to plate. Incubate 3-8 hours. Remove medium and replace with 6.0 ml DMEM-medium. Day 3: Change medium again to 6.0 mls of medium optimal for the cells to be infected. Move to  $32^\circ\text{C}$  either in the morning or afternoon depending on the Phoenix cell confluency and whether you will infect at 48 or 72 hrs after transfection. Day 4 or 5: Collect virus supernatant from transfected plates (6.0 ml) into 50 ml tubes and add protamine sulfate to a final concentration of 5  $\mu$ g/ml. Pass through a  $0.45\mu\text{m}$  filter. Count target cells and distribute  $10^7$  cells per 10 cm plate transfected to 50 ml tubes and pellet 5 min. Resuspend each pellet of target cells in virus supernatant and transfer to a 6 well plate at 1.0-1.2 ml per well. Seal plate with parafilm and centrifuge at RT for 30-90 min. at 2500 RPM. Remove parafilm and incubate plate over night at  $37^\circ\text{C}$ . Day 5: Collect and pellet each well of target cells. Resuspend in 3 ml medium and transfer back to the same 6well plate. Infection can be repeated by refeeding the Phoenix cells with 6ml fresh medium and reinfected the same cells again up to 3 times to increase % of cells infected (for instance at 48, 56, and 72 hours). Day 7 or Day 8: At 48 to 72 hrs. post infection, target cells are ready to analyze for expression.

This primary library will be used to infect at least  $10^9$  knock-in cells. After infection, the cells will be stimulated with IL-4 and two days later, peptide-containing cells (identified by the fluorescent reporter) that are negative for the knock-in reporter (i.e. where there is  $\epsilon$  promoter inhibition) will be sorted by FACS. This enriched, knock-in reporter negative population will be subjected to RT-PCR to amplify the integrated peptide sequences. The

PCR material will be used to construct a new “enriched” retrovirus peptide library to initiate the next screening round.

It will take approximately 5-7 rounds of enrichment to identify individual sequences capable of inhibiting the germline  $\epsilon$  promoter, as outlined below using an iterative screening equation.

$$R = \frac{V}{e^p + (Q + \sum_{i=0}^{\infty} \beta (1 + e)^p) + v}$$

The above equation mathematically models screening efficiency and provides a guideline for monitoring enrichment for inhibitory peptides.  $R$  = ratio of true positive cells over the total number of cells screened per round of selection;  $v$  = frequency of true positive cells (ie. # of cells expressing peptide inhibitors of IgE switch/synthesis);  $\epsilon$  = frequency of non-heritable false-positive cells (ie. # of cells in which IgE switch/synthesis is inhibited due to stimulation/screening inefficiencies, but are IgE positive in subsequent selection rounds);  $p$  = number of rounds of selection/enrichment applied to library screen;  $Q$  = initial frequency of cells with an heritable false-positive phenotype (ie. dominant-negative somatic mutation in cells that prevent IgE switch/synthesis);  $\beta$  = frequency of false-positives incurred by or during the selection/enrichment process.

Since we amplify enriched peptides by RT-PCR after each selection round, the equation can be simplified to

$$R = \frac{v}{e^p + Q + v}$$

By plugging in empirically-derived or estimated values for the variables, an estimate of how many selection rounds must be applied to a library before enrichment for IgE inhibitory peptide becomes apparent.

- 5 For the purposes of our screens, we engineer and select reporter cell lines in which the values of  $\epsilon$  and  $Q$  are low to minimize the number of screening rounds necessary to observe rare positive peptide "hits".

For example, IL-4 treatment upregulates the IgE switch reporter in 97% of cells, therefore  $\epsilon = 0.03$ . Of the uninduced cells, a second round of stimulation indicates that less than .01% of the starting population contain heritable false positives, therefore  $Q < 0.0001$ . A conservative estimate of IgE inhibitory peptides in the starting population is  $1/10^8$ , therefore  $v = 10^{-8}$ . Solving the equation for the number of selection rounds required to enrich to 50% true positive hits...

$$0.5 = \frac{10^{-8}}{(0.03)^p + 10^{-3} + 10^{-8}} \rightarrow p = 5 \text{ rounds}$$

15 The most important factor that influences the number of enrichment rounds necessary to identify individual peptide hits is the ratio between the real positive peptide hits in the original library and the heritable false positives. The frequency of real positive peptide hits is dependent upon the qualitative ability of the peptide to access and, in the correct conformation, bind to regulatory domains on proteins in the pathway of interest. Thus, preferably, multiple scaffolding structures are used for presentation of random peptide surfaces and also different localization sequences fused to those peptide structures. Enrichment of real positive peptides becomes less efficient with false positive rates above 2%. For this reason, great emphasis is placed on developing robust reporter constructs and cell lines.

Uneven RT-PCR amplification may decrease overall amplification of real peptides hits from one round to another. This is overcome by additional rounds of library enrichment and is why RT-PCR amplification is carefully monitored after each round of screening.

### Example 3

#### Screening for inhibitors of IgE secretion in cells that have already switched

After B cells have switched to production of IgE, there are several factors that determine when they will secrete IgE. By screening for peptide inhibitors of surface IgE expression, proteins that regulate IgE transcription, translation, assembly and trafficking may be identified.

The IgE<sup>+</sup> cell line, U266, expresses IgE on the surface and also secretes IgE. Antibodies against surface IgE heavy and light chains have been obtained and both are used to fluorescently mark IgE positive cells. The U266 line is consistently greater than 98.5% positive for membrane IgE.

Peptide library screening and target identification: The peptide library and enrichment protocols identical to those described in Example 2. As well, peptide hit validation and corresponding target protein identification will be performed as described in Example 2.

Development of an  $\epsilon$ -heavy chain GFP/BFP knock-in cell line derivative of U266: The cytoplasmic tail of the  $\epsilon$ -heavy chain in U266 cells will be engineered by homologous recombination to encode a GFP/BFP reporter as shown in Figure 8. This will produce a cell line that is fluorescent when  $\epsilon$ -heavy chains are produced. The GFP will be attached to the secretory exon to label secretory IgE heavy chains green. The BFP will be attached to the M2 exon to label membrane IgE heavy chains blue. This will allow discrimination of mRNA processing and translation between secretory versus membrane  $\epsilon$ -heavy chain transcripts.

The construct will contain an SV40 promoter-driven neomycin resistance gene which is flanked by loxP sites and inserted in the intron between the CH3 and CH4 exons (Figure 8). In addition, the HSV-TK gene will be cloned 3' of the longer homologous sequence region. U266 cells transfected with this construct will be selected for integration by culturing them in the presence of G418. The surviving cells will be cultured in ganciclovir to select against cells containing the HSV-TK gene (the HSV-TK gene at the 3' end will be deleted during the desired homologous recombination event). The remaining clones will be assessed for homologous recombination by PCR. Clones containing homologously-recombined constructs will be transfected with *cre* to mediate excision of the SV40 promoter/neomycin resistance gene in order to eliminate promoter interference. Excision will be verified by subdividing clones into parent and daughter pools and re-selecting the latter pool in G418. The parental cells corresponding to G418 sensitive daughter cells will be subdivided again and tested for GFP and BFP expression. Parental stocks of the most inducible clones will be used for subsequent screening.

#### Example 4

##### Development of an $\epsilon$ promoter GFP reporter cell line

The induction of the  $\epsilon$  promoter in response to IL-4/13 is the first recognizable step necessary for the switch to IgE. Blocking activation of this promoter should prevent B cells from switching to IgE. Inhibitors are predicted to interfere with IL-4/13 signaling as well as nuclear transcription of the  $\epsilon$  germline gene.

Three IgM<sup>+</sup>, EBV<sup>-</sup> human B cells lines (CA-46, MC116, and DND39; see Figure 4) that produce  $\epsilon$  germline transcripts in the presence of IL-4 will be infected with the following construct: a retroviral vector containing an IL-4 responsive 600 bp fragment of the  $\epsilon$  promoter in the reverse orientation followed by a splice site, GFP encoding sequence and a poly-adenylation sequence (Figure 10). Briefly, cells will be infected with the reporter construct and induced with IL-4. The cells will then be sorted by FACS for GFP reporter expression. The IL-4 will be removed and the cells will be sorted for the absence of reporter

fluorescence. From these sorts, several clones will be established that turn on the reporter in the presence of IL-4, indicating activation of the germline  $\epsilon$  promoter.

### Example 5

#### Screening of candidate agents using reporter cell line

The cell line of Example 4 is infected with a peptide library as described above. The peptide library is packaged into infectious viral particles (see Figure 7). The library is a mixture of random peptide sequences with and without a nuclear localization sequence (NLS) upstream of a reporter gene to identify infected cells and relative peptide expression (Figure 6).

Each screen will start with production of the primary retrovirus peptide library. This primary library will be used to infect at least  $10^9$   $\epsilon$  promoter reporter cells. After infection, the cells will be stimulated with IL-4 and two days later, the FACS will sort peptide-containing, reporter negative cells (i.e. where there is  $\epsilon$  promoter inhibition). This enriched, reporter negative population will be subjected to RT-PCR to amplify the integrated peptide sequences. The PCR material will be used to construct a new “enriched” retrovirus peptide library to initiate the next screening round.

It will take approximately 5-7 rounds of enrichment to identify individual sequences capable of inhibiting the germline  $\epsilon$  promoter (see discussion above regarding the statistics associated with enrichment). The most important factor that influences the number of enrichment rounds necessary to identify individual peptide hits is the ratio between real positive peptide hits in the original library and heritable false positives. The frequency of real positive peptide hits is dependent upon the qualitative ability of the peptide to get to and, in the correct conformation, bind to the regulatory domains on proteins in the pathway of interest. This is why we use multiple scaffolding structures for presentation of random peptide surfaces and also different localization sequences fused to those peptide structures (Appendix B).

Enrichment of real positive peptides becomes less efficient with false positive rates above

2%. For this reason, great effort is placed in developing robust reporter constructs and cell lines.

Once enrichment is achieved and individual peptide sequences are shown to effect inhibition of  $\epsilon$  promoter activation in an independent assay, they will be introduced into a standard set of secondary and orthogonal assays. Many of these assays will be performed in primary B cells to test the specificity and physiologic characteristics of the peptide inhibitor.

### Example 6

#### Generation of an $\epsilon$ promoter survival cell line.

Three different  $\text{IgM}^+$ ,  $\text{EBV}^-$  human B cells lines that produce  $\epsilon$  germline transcripts in the presence of IL-4 will be infected with a survival construct carrying a death gene and a drug selectable marker (Figure 10). Briefly, the retroviral construct consists of the 600 bp IL-4 inducible  $\epsilon$  promoter downstream of a self-inactivating (SIN) LTR, followed by a chimeric FAS receptor (FASr), the self-cleaving peptide 2a and, lastly, the drug-selectable puromycin resistance gene. The chimeric receptor is composed of the mouse FASr external domain and the human FASr transmembrane and cytoplasmic domains. A mouse specific anti-FASr antibody can be used which will bind only activated FASr produced by the survival construct. The 2a self-cleaving peptide allows equimolar amounts of the chimeric FASr and puromycin to be produced in the cell.

$\text{IgM}^+$  B cell lines infected with this construct in the presence of IL-4 will produce CD95, as well as puromycin resistance. Upon drug selection with puromycin, only cells containing IL-4 activated  $\epsilon$  promoters will survive. The remaining cells are infected with the peptide libraries and, when cultured in the presence of IL-4 and anti-FAS ( $\alpha\text{CD95}$ ) monoclonal antibodies, will express the chimeric FAS receptor and apoptose unless their  $\epsilon$  promoter has been blocked by a library peptide.

If problems arise due to over-expression of the chimeric FASr resulting in self-activation, other external domains will be used. We have already engineered a chimeric FASr containing the murine CD8 external domain as an alternative (Figure 10). If overexpression of the chimeric FASr results in self-activation, we have designed an alternative strategy in which the proposed construct contains the GFP gene in lieu of the puromycin resistance gene (Figure 10). Due to the mild transcriptional leakiness inherent to all SIN retroviral vectors, a small percentage of IgM<sup>+</sup> B cell clones infected with this construct will express low, detectable levels of GFP. These cells can be single-cell cloned by FACS, split into parent and daughter pools and tested for IL-4 inducible FASr expression-dependent apoptosis. Parent stocks of the most efficiently killed daughter cells will provide a continuous cell source for subsequent peptide screening assays. In addition, FASr ligation can be used to potentiate cell death and thus diminish background cell survival.

Additionally, IL-4 stimulation has been reported to diminish FAS-induced apoptosis in certain B-cell lines. To circumvent this potential difficulty, common suicide genes including Herpes Simplex Virus Thymidine Kinase (HSV-TK) or human cytochrome P450 2B1 in conjunction with ganciclovir or cyclophosphamide treatment, respectively, can replace FASr-mediated death (Figure 10). Alternatively, cell cycle arrest genes such as p21 can be used in place of toxic gene products (Figure 10). In this way, cells expressing peptides which prevent IL-4 induced overexpression of p21 will have a selective growth advantage and will quickly dominate the culture.

### Example 7

#### Screening in $\epsilon$ promoter survival cells

Using a peptide library generated as outlined above, the IgM<sup>+</sup> B cell lines described in Example 6 are infected with the survival construct. Leaky cells (constitutive expression of the  $\epsilon$  promoter) will be removed by incubation with the anti-mouse FASr antibody. Next, the cells are incubated in the presence of the inducer, IL-4, and the drug selection compound, puromycin. Cells that contain a construct that is inducible by IL-4 will be resistant and



survive. This produces a population with an exogenous  $\epsilon$  promoter that is IL-4 inducible. The peptide library is introduced into these cells and two days later they are induced with IL-4 in the presence of anti-mouse FASr monoclonal antibody. Cells carrying peptides that inhibit induction of the engineered  $\epsilon$  promoter fragment will not produce the chimeric FASr and will survive. After the survivors grow out (approximately 1 week), they will again be subjected to IL-4 and the anti-FASr treatment. The genes encoding the peptides responsible for the survivors will be rescued by RT-PCR and used to generate an enriched retroviral library. The identification of individual inhibitory peptides should occur in only 3-4 rounds since the false positive background for survival screens is lower than for FACS-based screening. Once enrichment is achieved and individual peptide sequences are independently shown to inhibit  $\epsilon$  promoter activation, these sequences will be introduced into a standard set of secondary and orthogonal assays.

#### Example 8

##### One-hybrid screens for identification of proteins that bind to switch $\epsilon$ region

Recombinase proteins that bind to the S $\epsilon$  region mediate the DNA rearrangement that generates a functional  $\epsilon$  heavy chain. They may be specific for  $\epsilon$  switching cells or may bind to other proteins that target them specifically to the S $\epsilon$  region. Breakpoints in the recombination of the switch  $\epsilon$  region to the switch  $\mu$  region occur in a limited area of the switch  $\epsilon$  region. Two stretches of the switch  $\epsilon$  region spanning the majority of breakpoints will be used as bait in a one-hybrid screen (Figure 2b). The cDNA libraries to be used are derived from the IgE positive cell line U266 (the assumption here is that the U266 line still contains the switch recombinase; certainly, the recombinase is turned off in plasma cells) and from human peripheral blood lymphocytes stimulated *in vitro* to switch with a high frequency to IgE.

The screening is summarized in Figure 3. The methods are as follows: Two stretches of the switch  $\epsilon$  region were cloned (Figure 2A) into EcoR I/Xba I sites of pHISi-1 (Clontech) to

5

10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044

155

20